1979

Characterizations of strong ergodicity for continuous time Markov chains

Mark Scott
Iowa State University

8000171

SCOTT, MARK

CHARACTERIZATIONS OF STRONG ERGODICITY FOR CONTINUOUS TIME MARKOV CHAINS.

IOWA STATE UNIVERSITY, PH.D., 1979

University Microfilms International 300 N. ZEEB ROAD, ANN ARBOR, MI 48106

Characterizations of strong ergodicity for

continuous time Markov chains

by

Mark Scott

A Dissertation Submitted to the

Graduate Faculty in Partial Fulfillment of

The Requirements for the Degree of

DOCTOR OF PHILOSOPHY

Major: Statistics

Approved:

Signature was redacted for privacy.

Signature was redacted for privacy.

In Charge of Major Work

Signature was redacted for privacy.

For the Major Department

Signature was redacted for privacy.

For the Graduate College

Iowa State University
Ames, Iowa

1979

TABLE OF CONTENTS

I. INTRODUCTION, REVIEW AND DEFINITIONS

A. Introduction and Review of Literature

In the study of Markov chains ergodic properties or loss of memory properties have received considerable attention. In the case of nonhomogeneous chains two distinct kinds of ergodic behavior must be distinguished, weak ergodicity, lack of memory, and strong ergodicity, convergence to a limiting distribution. Hajnal (1956) introduced the concepts of weak and strong ergodicity using ideas from Dobrushin (1956a). Dobrushin's ergodic coefficient, to be defined later, proved to be a convenient tool in this study.

For finite, homogeneous, discrete or continuous time Markov chains the concepts of ergodicity, strong ergodicity, and weak ergodicity coincide. But, as will be shown later, for infinite chains, the notions of ergodicity and strong ergodicity are separated.

This dissertation establishes new results in the characterization of strong ergodicity for continuous time Markov chains. In Chapters II, III, and IV we deal with homogeneous chains, while in Chapters VI and VII nonhomogeneous chains are examined.

For finite homogeneous ergodic Markov chains the eigenvalues of the transition matrix can be used to determine the geometric rate at which the chain converges. Isaacson and Luecke (1978) showed for infinite Markov chains that the spectrum plays a similar role. They showed that a discrete time chain is strongly ergodic if and only if the spectral radius of P-L is less than 1. Here P is the transition matrix of the chain and L is the row constant stochastic matrix to which P^n converges. They also showed that the "best" possible rate of convergence is given by the spectral radius of P-L. In Chapter II, we show that the results of Isaacson and Luecke remain valid for continuous chains.

In Chapter III we introduce the intensity of passage matrix, Q. Tweedie (1975) gave conditions on a Q-matrix which guaranteed that there exists a unique Markov chain with Q as its intensity matrix. In such a case the chain is said to be regular. The unique Markov chain determined by Q is called the Feller process. Tweedie also gave conditions on Q which imply ergodicity and recurrence for the Markov chain. Isaacson and Arnold (1978), with certain conditions on Q, showed that the "embedded" chain determined by Q is Cesaro strongly ergodic if and only if the original chain is strongly ergodic. Yong (1976) assumed conditions on Q, similar to those of Isaacson and Arnold, which permit him to define a related discrete time homogeneous Markov chain. He assumed that $\sup_i\{|q_{ii}|\} = q<+\infty$ and defined the chain $\overline{X}(n)$, generated by the powers of $\overline{P} = I + \frac{Q}{c}$, where

$c>q$. Yong was able to show that $\overline{X}(n)$ is irreducible, aperiodic, recurrent, positive recurrent, null recurrent, or ergodic if and only if the continuous time chain possesses the corresponding property. Also when either chain is ergodic the limiting distributions coincide. We employ the results found in Chapter II and functional analytic techniques to show that if $\overline{X}(n)$ is strongly ergodic, then the continuous time chain, $X(t)$, is strongly ergodic. Also, the converse holds when we impose conditions on the choice of the number $c>q$ and when we assume that the spectral radius of $\overline{P}-L$ is an element of the spectrum of $\overline{P}-L$. By assuming the above conditions, we are able to show that a rate of exponential convergence for $X(t)$ is an exponential function of the rate of geometric convergence for $\overline{X}(n)$.

In Chapter IV we strengthen the main result found in Chapter III by introducing mean visit times. Huang and Isaacson (1976) used mean visit times to characterize strong ergodicity for discrete time homogeneous and nonhomogeneous Markov chains. Isaacson and Arnold (1978) extended this work to the continuous time homogeneous case. Both works showed that a chain is strongly ergodic if and only if the mean visit times to some positive recurrent state are bounded. We shall use the above results to show that $X(t)$ is strongly ergodic if and only if $\overline{X}(n)$ is strongly ergodic. This will be done through the following argument. As mentioned in the previous paragraph, Yong showed that the mean visit time for some state is finite for X(t) if and only if the mean visit time is finite for $\overline{X}(n)$. He failed to mention the fact, which we prove, that there is a simple functional relationship between the mean visit times for X(t) and the mean visit times for $\overline{X}(n)$. Hence, we shall then be able to show the equivalence of strong ergodicity between the two chains.

Pitman (1974) used mean visit times to establish uniform rates of convergence of an ergodic discrete time Markov chain. He did not mention the possibility of extension to continuous time chains. In Chapter V we draw upon the results found in Chapters III and IV to determine the uniform rate of convergence for an ergodic continuous time chain.

In Chapter VI we begin the investigation of continuous time, nonhomogeneous Markov chains. The main result states that the chain is uniformly strongly ergodic if and only if the chain is uniformly ergodic and the mean visit times to some state are bounded over the starting states and starting times. The method of proof has the flavor of the proof given by Isaacson and Arnold for homogeneous chains. A major portion of this chapter establishes a uniform limit for a collection of discrete time nonhomogeneous chains induced by the continuous time chain, and a uniform bound for the mean visit times for the collection of discrete times chains is given.

The intensity of passage matrix for a nonhomogeneous chain depends on the time of movement. In Chapter VII we consider the case where $Q(s) = h(s)\cdot Q$, where Q is an intensity matrix and $h(s)$ satisfies certain regularity conditions. With these assumptions, we may use the results of Yong (1976), to show that strong ergodicity for the continuous time chain may be determined from the behavior of the discrete time chain determined by Q. In this special case we show that the limiting distribution L satisfies $LP(s,t) = L$ for all $s \leq t$.

B. Definitions and Notation

This dissertation will be concerned with viewing stochastic processes indexed by the nonnegative integers or the nonnegative reals. Therefore, it is convenient to denote the index set by T. We will be dealing with a special type of stochastic process, a Markov chain.

Definition 1.B.1:

A stochastic process $\{X(t): t \varepsilon T\}$ (or $X(t)$) with state space $S = \{0,1,2,\ldots\}$ is said to satisfy the Markov property if for any set of times $t_0 < t_1 < t_2 < \ldots < t_n < t$ and for any set of states $i_0, i_1, \ldots i_n, j$ it is true that

$$P(X(t) = j | X(t_n) = i_n,\ X(t_{n-1}) = i_{n-1}, \ldots X(t_0) = i_0)$$
$$= P(X(t) = j | X(t_n) = i_n).$$

This describes the fact that where the process moves on its next step depends only on what state it currently occupies. We will denote $P(X(t) = j | X(s) = i)$ by $p_{ij}(s,t)$. Thus $p_{ij}(s,t)$ represents the probability of occupying state j at time t given that at time s the process occupied state i. Thus given any $s<t$ we may define an infinite matrix $P(s,t)$, where

$$P(s,t) = \begin{bmatrix} p_{00}(s,t) & p_{01}(s,t) & p_{02}(s,t) & \cdots \\ p_{10}(s,t) & p_{11}(s,t) & p_{12}(s,t) & \cdots \\ p_{20}(s,t) & p_{21}(s,t) & p_{22}(s,t) & \cdots \\ \vdots & \vdots & \vdots & \vdots \\ p_{n0}(s,t) & p_{n1}(s,t) & p_{n2}(s,t) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

The elements of $P(s,t)$ satisfy the following conditions:

A) $p_{ij}(s,t) \geq 0$ for all $i,j \varepsilon S$ and all $s \leq t$.

B) $\sum_{j=0}^{\infty} p_{ij}(s,t) = 1$ for all $i \varepsilon S$ and all $s \leq t$.

C) For all times s, u, and t such that $s<u<t$ and for all i and j elements of S the Chapman-Kolmogorov equations hold:

$$p_{ij}(s,t) = \sum_{k=0}^{\infty} p_{ik}(s,u)\; p_{kj}(u,t).$$

D) $$\lim_{t \to s^+} p_{ij}(s,t) = \delta_{ij} = \begin{cases} 1 & \text{if } i=j \\ 0 & \text{if } i \neq j \end{cases}$$

Condition B gives the result that the sum of any row of $P(s,t)$ is equal to 1. Coupled with condition A we have, for any row of $P(s,t)$, a probability distribution on S. Condition C says that if one wants to examine the probability of going from state i at time s to state j at time t then one only need to examine an intermediate time u and all paths which lead from i to another state in u units of time and all paths which go from the intermediate state to state j in t-u units of time. Note that condition C also gives the following matrix equation. For all $s<u<t$,

$$P(s,t) = P(s,u) \cdot P(u,t).$$

If a stochastic process $X(t)$ satisfies the Markov property then $X(t)$ is said to be a Markov chain. Since T is either the set of nonnegative reals or nonnegative integers, we can immediately distinguish between two types of chains.

Definition 1.B.2:

Let $\{X(t):t\varepsilon T\}$ be a Markov chain. If T is the set of nonnegative integers then $\{X(k):k = 0,1,2...\}$ is said to be a discrete time Markov chain. If the set T is the nonnegative real numbers then $\{X(t):t>0\}$ is said to be a continuous time Markov chain.

We may further distinguish between types of chains by examining the behavior of the transition probabilities.

Definition 1.B.3:

A Markov chain $X(t)$ (discrete or continuous) is said to be time homogeneous (or homogeneous) provided

$$p_{ij}(t,t+h) = p_{ij}(0,h)$$

for all h and t elements of T and for all i and j elements of S. A Markov chain is said to be nonhomogeneous if the above condition fails to hold.

If a Markov chain is homogeneous then the probability of moving from one state to another in a fixed unit of time is independent of the starting time. Hence, we will write, for $X(t)$ homogeneous,

$$P(X(t) = j|X(s) = i) = p_{ij}(t-s).$$

Example 2.B.1:

Let $\{X(n):n = 0,1,2,\ldots\}$ be a Markov chain having probability transition matrix from time n to time n+1 of

$P(n,n+1) = \begin{bmatrix} .8 & .2 \\ .7 & .3 \end{bmatrix} = P$ for $n \geq 0$. Then, $X(n)$ is a discrete time homogeneous Markov chain, since the elements of P(n,n+1) do not depend on n.

Example 2.B.2:

Let $\{Y(n): n = 0,1,2,...\}$ be a Markov chain with probability transition matrix from time n to n+1 of

$$P(n,n+1) = \begin{bmatrix} .8 + \frac{1}{n+6} & .2 - \frac{1}{n+6} \\ .7 - \frac{1}{n+6} & .3 + \frac{1}{n+6} \end{bmatrix}, \quad \text{for } n \geq 0.$$

then, Y(n) is a discrete time nonhomogeneous Markov chain.

It may happen, that after a length of time, the chain is in some state or class of states from which it cannot return. We resolve this problem by assuming that any homogeneous Markov chain we examine has the following property.

Definition 1.B.4:

A homogeneous Markov chain X(t) is irreducible if for every pair of states i and j there exist times s and t such that $p_{ij}(s)>0$ and $p_{ji}(t)>0$.

Example 1.B.3:

Let X(n) be a discrete time homogeneous Markov chain with transition matrix

$$P = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

It is clear that $p_{00}(2) = 1$, $p_{11}(2) = 1$, $p_{01}(1) = 1$, and $p_{10}(1) = 1$ so that the chain is irreducible.

In the above example the chain determined by P was irreducible. Yet, the probability of returning to a particular state is positive only when the time of returning is a multiple of 2. This brings us to the next definition.

Definition 1.B.5:

Let $X(n)$ be a discrete time homogeneous Markov chain. State j is said to have period d if the following conditions hold:

A) $p_{jj}(n) = 0$, unless $n = md$ for some positive integer m.

B) d is the largest integer that has property A.

If d=1 the chain is said to be aperiodic.

In the previous example, we can see that both states have period 2, since $p_{jj}(2n) = 1$ for all n.

We are not only interested in the transition probabilities, but we would like to know the probability of a first visit at a particular time.

Definition 1.B.6:

If $X(n)$ is a discrete time homogeneous Markov chain, let $f_{ij}(n)$ denote the probability that the first visit to state j from state i occurs at time n. That is

$$f_{ij}(n) = P(X(n) = j,\ X(n-1)\neq j,\ldots X(1)\neq j \mid X(0) = i).$$

The analogous definition for continuous time homogeneous chains is found in the following manner. If $X(t)$ is a continuous time homogeneous Markov chain the waiting time to move from state i is given by the following random variable

$$W_i = \inf\{t:t>0,\ X(t)\neq i\}.$$

Let T_{ij} be the random variable which measures the time until the first visit from state i to state j. That is

$$T_{ij} = \inf\{t:t>W_i,\ X(t) = j\}.$$

With this definition we have the following definition.

Definition 1.B.7:

The distribution function $F_{ij}(t) = P(T_{ij}\leq t)$ represents the first visit distribution from state i to state j.

Definition 1.B.8:

If $X(t)$ is a homogeneous Markov chain then,

i) $f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n)$ = probability of ever visiting state j from state i, when the chain is discrete.

ii) $F_{ij} = \int_0^{\infty} dF_{ij}(t)$ = probability of ever visiting state j from state i, when the chain is continuous.

Using the above definition, we can characterize the individual states of the chain by the behavior of f_{ij} or F_{ij}.

Definition 1.B.9:

A state i is said to be recurrent provided

i) $f_{ii} = 1$ when the chain is discrete.

ii) $F_{ii} = 1$ when the chain is continuous.

If $f_{ii}<1$ ($F_{ii}<1$) then state i is said to be transient.

If $f_{ii} = 1$ then the first return probabilities form a probability distribution on the state space. Hence we may talk about the mean return time.

Definition 1.B.10:

If $f_{ii} = 1$ then define

$$m_{ii} = \sum_{n=1}^{\infty} n.f_{ii}(n)$$

to be the mean return time for state i. If $m_{ii}=\infty$ then state i is said to be null recurrent. If $m_{ii}<\infty$ then state i is said to be positive recurrent.

Similarly if $F_{ii}=1$, then $F_{ii}(t)$ is a distribution function of the proper random variable T_{ii}. Hence, the expectation of T_{ii} makes sense.

Definition 1.B.11:

If $F_{ii}=1$ then define

$$m_{ii} = E(T_{ii}) = \int_0^\infty t dF_{ii}(t)$$

to be the mean return time for state i. If $m_{ii}<\infty$, state i is positive recurrent, otherwise it is null recurrent.

If X(t) is an irreducible Markov chain, then if one state is recurrent, $f_{ii}=1$ or $F_{ii}=1$, then all states are recurrent. Hence, if one state in an irreducible chain is recurrent then $f_{ij}=1$ for all i and j when the chain is discrete, or $F_{ij}=1$ for all i and j if the chain is continuous. Thus we may define, for a recurrent state j, the mean visit time from i to j.

Definition 1.B.12:

If state j is recurrent then

i) $m_{ij} = \sum_{n=1}^{\infty} n.f_{ij}(n)$ = mean visit time from i to j for a discrete time chain

ii) $m_{ij} = \int_0^\infty t \cdot dF_{ij}(t)$ = mean visit time from i to j for a continuous time chain.

It is well-known that given the transition probabilities and an initial distribution the unconditional distribution of the chain may be determined. After observing the chain after a long period of time we would like to see if the probability of being in a particular state is independent of the starting state. This occurrence falls under the general heading of ergodic behavior.

Definition 1.B.13:

Let $X(t)$ be a homogeneous Markov chain. If

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j$$

independent of i for all j and $\sum_{j=0}^{\infty} \pi_j = 1$, $\pi_j \geq 0$, then we say that the chain is ergodic.

The vector $\pi = (\pi_0, \pi_1, \ldots)$ is called the stationary distribution for $X(t)$, and satisfies

$$\pi P(t) = \pi$$

for all times t.

Example 1.B.4:

Consider the continuous time homogeneous Markov chain $X(t)$ with transition probability matrix

$$P(t) = \begin{pmatrix} e^{-t} & 1-e^{-t} \\ 0 & 1 \end{pmatrix}.$$

Clearly $\lim_{t\to\infty} p_{00}(t) = p_{10}(t) = 0$ and $\lim_{t\to\infty} p_{01}(t) = p_{11}(t) = 1$ so that the chain is ergodic.

For a nonhomogeneous chain ergodicity is defined in a similar manner. That is, we require, for each s,

$$\lim_{t\to\infty} p_{ij}(s,t) = \pi_j .$$

If the convergence is uniform with respect to s then we say the chain is uniformly ergodic.

Ergodicity guarantees the convergence of the elements of P(s,t). We may want to see if the matrix P(s,t) converges, in some sense. Before we do this, we need the following definition.

<u>Definition 1.B.14</u>:

If $p = (p_0, p_1, \ldots)$ is a vector, then define the norm of p as

$$||p|| = \sum_{i=0}^{\infty} |p_i| .$$

If $A = (a_{ij})$ is a matrix, define the norm of A by

$$||A|| = \sup_i \sum_{j=0}^{\infty} |a_{ij}| .$$

It is easy to show that $||\cdot||$ is actually a norm. We will use this norm to discuss the convergence of P(s,t).

Definition 1.B.15:

A Markov chain $X(t)$ is said to be strongly ergodic if there exists a row constant stochastic matrix L, such that for all times s

$$\lim_{t\to\infty} ||P(s,t)-L|| = 0.$$

Suppose $X(t)$ is strongly ergodic. Let $\pi = (\pi_0,\pi_1,\ldots)$ be any row of L. For any times s and t

$$|p_{ij}(s,t)-\pi_j| \le \sum_{k=0}^{\infty} |p_{ik}(s,t)-\pi_k| \le \sup_i \sum_{k=0}^{\infty} |p_{ik}(s,t)-\pi_k|$$
$$= ||P(s,t)-L||.$$

Taking limits, $\lim_{t\to\infty} p_{ij}(s,t) = \pi_j$ for each s. Hence $X(t)$ is ergodic.

If $X(t)$ is a strongly ergodic homogeneous Markov chain then each row of L is π, the stationary distribution for $P(t)$. If $X(t)$ is a strongly ergodic nonhomogeneous Markov chain and $||P(s,t)-L||$ goes to zero uniformly in s as t goes to infinity then we say that $X(t)$ is uniformly strongly ergodic.

For homogeneous, aperiodic, irreducible, finite state, discrete time Markov chains the concepts of ergodicity and strong ergodicity coincide. Yet, in the infinite state space situation we may have elementwise convergence without having matrix convergence.

Example 1.B.5:

Let X(n) be the discrete time homogeneous Markov chain determined by the stochastic matrix

$$P = \begin{bmatrix} q & p & 0 & 0 & 0 & \cdots \\ q & 0 & p & 0 & 0 & \cdots \\ 0 & q & 0 & p & 0 & \cdots \\ 0 & 0 & q & 0 & p & \cdots \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \vdots & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \end{bmatrix}$$

Here $p+q = 1$ and $0<p<q<1$. This is a random walk with an elastic barrier. For any two states i and j with $j>i$ we can see that $p_{ij}(j-i) \geq p^{j-i}>0$. If $j<i$ then $p_{ji}(i-j) \geq q^{i-j} > 0$. Hence the chain is irreducible. Since $p_{00}(1) = q>0$ state 0 is aperiodic. Hence all states are aperiodic.

To show that the chain is recurrent, we need the following result.

Theorem 1.B.1:

Let X(n) be an irreducible, aperiodic, homogeneous, discrete time Markov chain. Consider the system of linear equations

$$\pi_j = \sum_{i=1}^{\infty} \pi_i p_{ij}(1), \quad j = 0,1,2,\ldots .$$

Then all states are recurrent and, in fact, positive recurrent if and only if there exists a solution π with $\pi_j \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$.

Proof: (See Cinlar (1975, p. 175))

The solution of $\pi P(1) = \pi$ for this example is given by following system of linear equations:

$$\pi_0 \cdot q + \pi_1 \cdot p = \pi_0$$
$$\pi_0 p + \pi_2 q = \pi_1$$
$$\pi_1 p + \pi_3 q = \pi_2$$
$$\vdots$$
$$\pi_{n-1} p + \pi_{n+1} q = \pi_n$$
$$\vdots$$

and $\sum_{j=0}^{\infty} \pi_j = 1$. Clearly

$$\pi_1 = \frac{1-q}{q}\pi_0 = \frac{p}{q}\pi_0$$

$$\pi_2 = \frac{1}{q}(\pi_1 - p\pi_0) = \frac{1}{q}\left(\frac{p}{q} - 1\right)\pi_0 = \frac{p^2}{q^2}\pi_0$$

$$\pi_3 = \frac{1}{q}(\pi_2 - p\pi_1) = \frac{1}{q}\left(\frac{p^2}{q^2} - \frac{p^2}{q}\right)\pi_0 = \frac{p^3}{q^3}\pi_0$$
$$\vdots$$
$$\pi_n = \frac{p^n}{q^n}\pi_0$$
$$\vdots$$

Hence

$$1 = \sum_{n=0}^{\infty} \pi_n = \sum_{n=0}^{\infty} \left(\frac{p}{q}\right)^n \pi_0 = \frac{q}{q-p}\,\pi_0 .$$

$$\pi_0 = \frac{q-p}{q},\ \pi_1 = \frac{p}{q^2}\,(q-p),\ \ldots\ \pi_n = \frac{p^n}{q^{n+1}}\,(q-p),\ \ldots .$$

By the theorem the chain is recurrent and hence all states are positive recurrent. Having all states positive recurrent is sufficient for ergodicity. So

$$\lim_{n\to\infty} p_{ij}(n) = \pi_j = \frac{p^j}{q^{j+1}}(q-p) \quad \text{for} \quad j = 0,1,2,\ldots .$$

The chain, however, is not strongly ergodic. To see this, we use the following theorem.

<u>Theorem 1.B.12</u>:

A discrete time, homogeneous, aperiodic, irreducible Markov chain is strongly ergodic if and only if $\delta(P(n))<1$ for some n. Where

$$\delta(P(n)) = \frac{1}{2} \sup_{i,j} \sum_{k=0}^{\infty} |p_{ik}(n) - p_{jk}(n)| .$$

For this example, P(n) is found by finding P^n, and

$$P^n = \begin{bmatrix} \ddots & & \phi \\ & & \ddots \\ \phi & & \end{bmatrix} .$$

Hence we can find, for any n, 2 rows, say i and k, such that the ith row is

$$(p_{i0}(n),\ p_{i1}(n),\ \ldots\ p_{ij}(n),\ 0,\ 0,\ 0,\ldots)$$

and the kth row is

$$(0,\ 0,\ 0,\ \ldots,\ 0,\ p_{kh}(n),\ p_{k,h+1}(n),\ \ldots) \text{ where } h>j.$$

Hence

$$\sum_{m=0}^{\infty} |p_{im}(n)-p_{km}(n)| = \sum_{m=0}^{j} p_{im}(n) + \sum_{m=h}^{\infty} p_{km}(n) = 1+1=2.$$

Thus,

$$\frac{1}{2} \sup_{i,k} \sum_{m=0}^{\infty} |p_{im}(n)-p_{km}(n)| = \delta(P(n)) = 1.$$

Since this is true for all n, the chain is not strongly ergodic.

Note: the number $\alpha(P(n)) = 1-\delta(P(n))$, where $\delta(P(n))$ was defined in Theorem 1.B.2, is Dobrushin's ergodic coefficient.

II. STRONG ERGODICITY FOR CONTINUOUS TIME HOMOGENEOUS CHAINS USING SPECTRAL CONDITIONS

For a strongly ergodic Markov chain, on a finite dimensional state space, determined by a stochastic matrix P, it is well-known that

$$\beta = \sup\{|\lambda|: \lambda \text{ is an eigenvalue of } P,\ |\lambda| \neq 1\}$$

is the best possible rate of convergence of $P(n) = P^n$ to L. In order to study infinite dimensional chains one needs to define the spectrum and spectral radius of P.

Definition 2.1:

If A is a matrix then the spectrum of A is denoted and defined by

$$\sigma(A) = \{\lambda: A-\lambda I \text{ does not have a continuous inverse}\}.$$

Definition 2.2:

$r(A) = \sup\{|\lambda|: \lambda \varepsilon \sigma(A)\}$ is called the spectral radius of A.

It can be shown that $\lim\limits_{n\to\infty} ||A^n||^{\frac{1}{n}} = r(A)$.

For an infinite dimensional, homogeneous Markov chain X(n), it has been shown, by Isaacson and Luecke (1978), that if X(n) is ergodic then X(n) is strongly ergodic if and only if $r(P-L)<1$. In this case $r(P-L) =$

$\sup\{|\lambda|: \lambda\varepsilon\sigma(P),\ \lambda\neq 1\}$ and $r(P-L)$ is the "best" geometric rate of convergence of $P(n)$ to L.

In this chapter we show the same results hold true for continuous time, homogeneous Markov chains. We first show that $X(t)$ is strongly ergodic if and only if $r(P(1)-L)<1$. Through a series of lemmas we are able to show

$$\begin{aligned} r(P(1)-L) &= \inf_t (\delta(P(t)))^{\frac{1}{t}} = \lim_{n\to\infty} (\delta(P(n)))^{\frac{1}{n}} \\ &= \lim_{t\to\infty} (\delta(P(t)))^{\frac{1}{t}} = \lim_{t\to\infty} ||P(t)-L||^{\frac{1}{t}}. \end{aligned}$$

We then show for strongly ergodic $X(t)$, the convergence of $P(t)$ to L is at an exponential rate. Although this result is by no means new, the rate given by the standard proof depends on an integer time. From the string of equalities we are able to improve on this rate.

For the remaining part of this chapter, $X(t)$ will be a continuous time, homogeneous, ergodic Markov chain. $P(t)$ will represent the matrix of transition probabilities of $X(t)$.

We begin by stating some well-known inequalities involving the δ-coefficient and the norm.

Lemma 2.1:

If P_1 and P_2 are stochastic matrices then

$\delta(P_1 \cdot P_2) \leq \delta(P_1) \cdot \delta(P_2)$.

Proof:

See Isaacson and Madsen (1976, p. 145).

Lemma 2.2:

If P is stochastic and R is a matrix such that

$||R|| < +\infty$ and $\sum_{k=0}^{\infty} r_{ik} = 0$ for all i, then

$||RP|| \leq ||R|| \cdot \delta(P)$.

Proof:

See Isaacson and Madsen (1976, p. 147).

Lemma 2.3:

For any matrices A and B

a) $||AB|| \leq ||A|| \cdot ||B||$

and

b) $||A+B|| \leq ||A||+||B||$.

If X(t) is strongly ergodic then there exists a real number T such that $\delta(P(T))<1$. For any $t>T$

$$P(t) = P(t-T+T) = P(t-T) \cdot P(T).$$

Hence by Lemma 2.1

$$\delta(P(t)) \leq \delta(P(t-T)) \cdot \delta(P(T)) \leq \delta(P(T)) < 1.$$

Note that $\delta(P(t))$ is a monotone decreasing function of t.

We can use the above fact to characterize strong ergodicity for X(t) in terms of strong ergodicity for the discrete time chain generated by the powers of $P(1)$.

Lemma 2.4:

$X(t)$ is strongly ergodic if and only if the discrete time chain generated by $P(n)$, n an integer, is strongly ergodic.

Proof:

If $X(t)$ is strongly ergodic then there exists a time t such that $\delta(P(t))<1$. Let n be any integer larger than t. By the comment after Lemma 2.3 $\delta(P(n)) \leq \delta(P(t)) < 1$. Hence the chain generated by $P(n)$ is strongly ergodic.

The converse is proved in a similar manner by reversing the roles of n and t.

For a discrete time homogeneous Markov chain, strong ergodicity is equivalent to having $r(P-L) < 1$. We now show this is true when discrete is replaced by continuous.

Theorem 2.1:

$X(t)$ is strongly ergodic if and only if $r(P(1)-L) < 1$.

Proof:

$X(t)$ is strongly ergodic if and only if $P(1)$ generates a strongly ergodic discrete time Markov chain. From the work of Isaacson and Luecke (1978) the second statement is equivalent to having $r(P(1)-L) < 1$.

We now begin the study of the rate that P(t) converges to L, when X(t) is strongly ergodic. For the discrete time, homogeneous, strongly ergodic Markov chain generated by the powers of P(1) we have, by Isaacson and Luecke, that $\lim_{n\to\infty}(\delta(P(n)))^{\frac{1}{n}}$ exists and equals $\inf_{n}(\delta(P(n)))^{\frac{1}{n}}$. Let [x] be the greatest integer less than or equal to x. By the argument after Lemma 2.3 we have for any t

$$\delta(P([t]+1)) \leq \delta(P(t)) \leq \delta(P([t]))$$

Thus

$$(\delta(P([t]+1)))^{\frac{1}{[t]+1}\cdot\frac{[t]+1}{t}} \leq (\delta(P(t)))^{\frac{1}{t}}$$

$$\leq (\delta(P([t])))^{\frac{1}{[t]}\cdot\frac{[t]}{t}}$$

As $t\to\infty$ $\frac{[t]}{t} \to 1$ and $\frac{[t]+1}{t} \to 1$. Also

$$\lim_{t\to\infty}(\delta(P([t]))^{\frac{1}{[t]}} = \lim_{n\to\infty}(\delta(P(n)))^{\frac{1}{n}}$$

where n is an integer. Hence

$$\lim_{t\to\infty}(\{\delta(P([t]+1))\}^{\frac{1}{[t]+1}})^{\frac{[t]+1}{t}} \leq \lim_{t\to\infty}[\delta(P(t))]^{\frac{1}{t}}$$

$$\leq \lim_{t\to\infty}(\{\delta(P([t]))\}^{\frac{1}{[t]}})^{\frac{[t]}{t}}$$

or

$$\lim_{n\to\infty}(\delta(P(n+1)))^{\frac{1}{n+1}} \leq \lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} \leq \lim_{n\to\infty}(\delta(P(n)))^{\frac{1}{n}}\ .$$

Therefore

$$\lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} \text{ exists and equals } \inf_n(\delta(P(n)))^{\frac{1}{n}}.$$

We are now ready to replace the $\inf_n(\delta(P(n)))^{\frac{1}{n}}$ by $\inf_t(\delta(P(t)))^{\frac{1}{t}}$.

Lemma 2.5:

If $X(t)$ is strongly ergodic then

$$\lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} = \inf_t(\delta(P(t)))^{\frac{1}{t}}.$$

Proof:

Let $\beta = \inf_t(\delta(P(t)))^{\frac{1}{t}}$, then there exists, for any $\varepsilon>0$, a real T such that $(\delta(P(T)))^{\frac{1}{T}} \leq \beta+\varepsilon$. Let t be any real and write $t = nT+r$, where n is an integer and $0\leq r\leq T$. consider

$$\begin{aligned}(\delta(P(t)))^{\frac{1}{t}} &= (\delta(P(nT+r)))^{\frac{1}{t}} \leq (\delta(P(nT)))^{\frac{1}{t}}\cdot(\delta(P(r)))^{\frac{1}{t}} \\ &\leq (\delta(P(T)))^{\frac{n}{t}} = (\delta(P(T)))^{\frac{1}{T}\cdot\frac{nT}{t}} \\ &\leq (\beta+\varepsilon)^{\frac{nT}{t}}.\end{aligned}$$

As $t\to\infty$, $nT/t\to 1$. Hence $\limsup_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} \leq \beta+\varepsilon$. Since this holds for all $\varepsilon>0$ and since $(\delta(P(t)))^{\frac{1}{t}} \geq \beta$ for all t we have that $\lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} = \inf_t(\delta(P(t)))^{\frac{1}{t}}$.

Combining this result with the comments preceding Lemma 2.5 we have

$$\lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}} = \lim_{n\to\infty}(\delta(P(n)))^{\frac{1}{n}} = \inf_t(\delta(P(t)))^{\frac{1}{t}}.$$

Again let $X(t)$ be strongly ergodic. Hence the chain determined by the powers of $P(1)$ is strongly ergodic. Hence $\lim_{n\to\infty}||P(n)-L||^{\frac{1}{n}}$ exists and equals $r(P(1)-L)$. We now replace the limit as n goes to infinity by the limit as t goes to infinity.

Lemma 2.6:

If $X(t)$ is strongly ergodic then $\lim_{t\to\infty}||P(t)-L||^{\frac{1}{t}}$ exists and equals $r(P(1)-L)$.

Proof:

From the comments given above $\lim_{n\to\infty}||P(n)-L||^{\frac{1}{n}} = r(P(1)-L)$. For any $s<t$

$$\begin{aligned} ||P(t)-L|| &= ||P(t-s+s)-L|| = ||P(t-s)\cdot P(s)-L|| \\ &= ||P(t-s)\cdot(P(s)-L)|| \\ &\le ||P(t-s)||\cdot||P(s)-L|| \\ &= ||P(s)-L||. \end{aligned}$$

Hence for any real t

$$||P([t]+1) - L|| \le ||P(t)-L|| \le ||P([t])-L||.$$

As in the argument preceding Lemma 2.5, it is easily shown that

$$\lim_{n\to\infty}||P(n+1)-L||^{\frac{1}{n+1}} \leq \lim_{t\to\infty}||P(t)-L||^{\frac{1}{t}} \leq \lim_{n\to\infty}||P(n)-L||^{\frac{1}{n}}.$$

Since $\lim_{n\to\infty}||P(n)-L||^{\frac{1}{n}} = r(P(1)-L)$, $\lim_{t\to\infty}||P(t)-L||^{\frac{1}{t}}$ exists and equals $r(P(1)-L)$. Thus the proof is complete.

Isaacson and Luecke also showed for a strongly ergodic discrete time Markov chain $r(P-L) = \inf_n(\delta(P(n)))^{\frac{1}{n}}$. Applying this to the strongly ergodic chain generated by $P(1)$ we have

$$\lim_{n\to\infty}(\delta(P(n)))^{\frac{1}{n}} = \inf_t(\delta(P(t)))^{\frac{1}{t}} = \lim_{t\to\infty}(\delta(P(t)))^{\frac{1}{t}}$$

$$= r(P(1)-L) = \lim_{t\to\infty}||P(t)-L||^{\frac{1}{t}}.$$

We now give the result concerning exponential convergence. Using the above equalities we can improve on the exponential rate.

Lemma 2.7:

If $X(t)$ is strongly ergodic then $P(t)$ converges at an exponential rate.

Proof:

Since $X(t)$ is strongly ergodic there exists a row constant stochastic matrix L such that

$$\lim_{t\to\infty} ||P(t)-L|| = 0$$

and $LP(t) = P(t)\cdot L = L^k = L$ for all real t and all integers $k \geq 1$. Since there exists real T_1, such that $\delta(P(T_1)) < 1$, define

$$T = \inf\{n: n \geq T_1 \text{ and } n \text{ is an integer}\}.$$

Since the δ-coefficient is monotone decreasing we have

$$\delta(P(T)) \leq \delta(P(T_1)) < 1.$$

For any real $t \geq T$, we may write $t = q\cdot T+r$ where $0 \leq r < T$ and $q = [\frac{t}{T}]$. Thus for $t \geq T$

$$\begin{aligned} ||P(t)-L|| &= ||P(Tq+r)-L|| = ||P(r)\cdot P(Tq)-L|| \\ &= ||P(r)\cdot P(Tq)-L\cdot P(Tq)|| \\ &\leq ||(P(r)-L)\cdot P(Tq)||. \end{aligned}$$

The matrix $P(r)-L$ has row sums equal to zero, and $||P(r)-L|| \leq 2$. Hence by Lemma 2.2 $||P(t)-L|| \leq 2\cdot\delta(P(Tq))$. Since q is an integer $P(qT) = P(T+T\ldots+T) = [P(T)]^q$, so that

$$\delta(P(qT)) \leq (\delta(P(T)))^q.$$

Put $\eta = \delta(P(T)) < 1$. Hence $||P(t)-L|| \leq 2.\eta^q$. Since $q = [\frac{t}{T}], q \geq \frac{t}{T} - 1$. Thus

$$\eta^q \leq \eta^{\frac{t}{T}-1} = \frac{(\eta^{\frac{1}{T}})^t}{\eta}. \quad \text{Hence}$$

$$||P(t)-L|| \le 2\cdot\eta^q \le \frac{(\eta^{\frac{1}{T}})^t}{\eta}\cdot 2.$$

Put $c = \frac{2}{\eta}$ and $\gamma = \eta^{\frac{1}{T}} < 1$. Thus

$$||P(t)-L|| \le c.\gamma^t$$

for $t \ge T$.

Now if $t<T$ then $||P(t)-L|| \le 2 = 2\cdot\frac{\gamma^t}{\gamma^t}$. Since $t<T$ and $\gamma = \eta^{\frac{1}{T}} < 1$, we have $\gamma^t > \gamma^T$ or $\gamma^{-t} < \gamma^{-T}$. Thus $||P(t)-L|| \le 2\cdot\gamma^t\cdot\gamma^{-t} \le 2\cdot\gamma^t\cdot\gamma^{-T} = \frac{2}{\eta}\cdot\gamma^t = c.\gamma^t$. Hence for all t $||P(t)-L|| \le c.\gamma^t$ so $P(t)$ converges to L at an exponential rate.

The proof of Lemma 2.7 hinged on choosing an integer time T such that $(\delta(P(T)))^{\frac{1}{T}} < 1$. From the equalities preceding Lemma 2.7 we have $(\delta(P(T)))^{\frac{1}{T}} \ge \inf_t(\delta(P(t)))^{\frac{1}{t}} = r(P(1)-L) = \lim_{t\to\infty}||(P(t)-L)||^{\frac{1}{t}}$. Let $\beta = r(P(1)-L)$ and define $\alpha_t = \beta - ||P(t)-L||^{\frac{1}{t}}$. Then $\lim_{t\to\infty}\alpha_t = 0$ and $||P(t)-L|| = (\beta+\alpha_t)^t$ for all t. Hence the "best" rate of convergence is given by $\beta = r(P(1)-L)$ and the other rate, $\delta(P(T))^{\frac{1}{T}}$, will never be smaller than β since $\beta = \inf_t(\delta(P(t)))^{\frac{1}{t}}$.

III. STRONG ERGODICITY FOR CONTINUOUS TIME HOMOGENEOUS CHAINS USING A RELATED DISCRETE TIME CHAIN

In practice one does not deal with the transition functions $P(t)$. If one is modelling some phenomena, by a continuous time Markov chain, the transition intensities are estimated from the data first.

<u>Definition 3.1</u>:

Let $P(t)$ be the transition probabilities for a continuous time Markov chain $X(t)$. Define

a) $$q_{ii} = \lim_{h\to 0^+} \frac{p_{ii}(h)-1}{h}$$

and

b) $$q_{ij} = \lim_{h\to 0^+} \frac{p_{ij}(h)}{h} \quad \text{for} \quad i\neq j.$$

We call q_{ii} the intensity of passage out of state i. Similarly, for $i\neq j$ q_{ij} is called the intensity of passage from state i to state j.

It can be shown that q_{ii} is always nonpositive and may take on the value $-\infty$. Also q_{ij} exists and is finite for all $i\neq j$.

The reasoning behind the naming "intensity of passage" may be found via the following argument. If $i=j$, then

$$1-p_{ii}(h) = P(X(h) \neq i \mid X(0)=i).$$

From the definition of the derivative, for small h we have

$$P(X(h)\neq i\,|\,X(0)=i) = -q_{ii}\cdot h+o(h).$$

Where $o(h)$ is a term such that $\frac{o(h)}{h}\to 0$ as $h\to 0$. Similarly for $i\neq j$ $p_{ij}(h) = P(X(h)=j\,|\,X(0)=i)$. Thus for small values of h

$$P(X(h) = j\,|\,X(0)=i) = q_{ij}\cdot h+o(h).$$

Let $Q = (q_{ij})$ be the matrix of transition intensities. Note that each row of Q satisfies $-q_{ii} = \sum_{j\neq i} q_{ij}$.

It can be shown that Q satisfies the matrix differential equation $P'(t) = P(t)\cdot Q = QP(t)$ (Reuter and Ledermann (1954)). Yet the general equation $F'(t) = F(t)\cdot Q = QF(t)$ may have solutions other than $P(t)$. Reuter and Ledermann gave a sufficient condition for the solution of $F'(t) = F(t)Q = QF(t)$ to be unique. If we assume $\sup_i\{|q_{ii}|\}=q<\infty$, then the solution of $F'(t) = F(t)\cdot Q = QF(t)$ will be unique and the solution is given by $P(t)$.

Hence assume that $\sup_i\{|q_{ii}|\} = q<+\infty$. Then $P'(t) = Q\cdot P(t) = P(t)\cdot Q$. Define $F(t) = I + \sum_{j=1}^{\infty} \frac{Q^j\cdot t^j}{j!}$. It is easily shown that $F'(t) = Q\cdot F(t) = F(t)\cdot Q$. Hence $P(t) = F(t)$ so that we may write $P(t) = I + \sum_{j=1}^{\infty} \frac{Q^j t^j}{j!} = \exp(Qt)$. Here $\exp(Qt)$ is regarded as an infinite series of matrices.

Since Q is more likely to be available than $P(t)$, one would like to impose conditions on a Q which guarantee ergodic properties of X(t). One method is to look at a discrete time chain determined by Q. Define a matrix R as follows,

$$r_{ij} = \begin{cases} q_{ij}/|q_{ii}| & \text{for } i\neq j,\ q_{ii}\neq 0 \\ 0 & \text{for } i=j \text{ and } q_{ii}\neq 0 \\ 1 & \text{for } i=j \text{ and } q_{ii}=0 \end{cases}$$

<u>Example 3.1</u>:

Let

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & -2 \end{pmatrix}$$

then

$$R = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}.$$

It is clear that R is a stochastic matrix, hence the powers of R generate a discrete time Markov chain, Y(n). Y(n) is called the "embedded" chain determined by Q.

For an irreducible, continuous time Markov chain $-q_{ii}>0$. If not, then $q_{ii}=0$ for some i then we must have $p_{ii}(t) = 1$ for all t. Hence the chain would not be irreducible. In terms of Y(n), this says that even though X(t) is irreducible, Y(n) may not be aperiodic.

Example 3.2:

If $Q = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}$ then $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. The chain determined by Q is strongly ergodic. Yet Y(n) is a periodic chain, with both states having period 2. Therefore Y(n) cannot be strongly ergodic.

To eliminate the problem of periodicity of Y(n), we shall define a different discrete time chain determined by Q. Since $\sup_i\{|q_{ii}|\} = q<+\infty$, let $c>q$ and define $\overline{P} = I + \frac{Q}{c}$, (see Yong (1976)).

Lemma 3.1:

$\overline{P}$ generates a discrete time Markov chain, $\overline{X}(n)$.

Proof:

For any i and j $\overline{p}_{ij} = \delta_{ij} + \frac{q_{ij}}{c}$. If i=j then $\overline{p}_{ii} = 1 + \frac{q_{ii}}{c}$. Since $c>|q_{ii}|$, for all i we have that

$$-1 \leq \frac{q_{ii}}{c} \leq 0$$

or

$$0 \leq \overline{p}_{ii} \leq 1.$$

Similarly,

$0 \le \bar{p}_{ij} \le 1$. Consider

$$\sum_{j=0}^{\infty} \bar{p}_{ij} = \sum_{j=0}^{\infty} (\delta_{ij} + \frac{q_{ij}}{c}) = 1 + \frac{1}{c} \cdot \sum_{j=0}^{\infty} q_{ij}$$

$$= 1 + \frac{1}{c}(q_{ii} + \sum_{j \neq i} q_{ij}) = 1 + \frac{1}{c}(q_{ii} - q_{ii}) = 1.$$

Hence $\bar{P}$ is stochastic for each $c>q$. Thus by taking integer powers of $\bar{P}$, we generate a discrete time chain $\bar{X}(n)$.

Example 3.3:

Suppose

$$P(t) = \begin{pmatrix} .01 + .99 \exp(-100t) & .99 - .99 \exp(-100t) \\ .01 - .01 \exp(-100t) & .99 + .01 \exp(-100t) \end{pmatrix}$$

Then

$$\lim_{t \to 0^+} \frac{P(t) - I}{t} = Q = \begin{pmatrix} -99 & 99 \\ 1 & -1 \end{pmatrix}.$$

Hence $R = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. With $c = 198$ we have $\bar{P} = \begin{pmatrix} 1/2 & 1/2 \\ 1/198 & 197/198 \end{pmatrix}$.

$X(t)$ is never periodic while the chain determined by R is periodic. Yet $\bar{X}(n)$ is aperiodic. More importantly $\bar{P}$ reflects the staying power of the chain $X(t)$ while the embedded chain affords no such interpretation. That is, since $-q_{00} = 99$ and $-q_{11} = 1$ the chain moves faster out of state 0 than state 1, having $\overline{p}_{01} = \frac{1}{2}$ and $\overline{p}_{10} = 1/198$ reflects this property. Yet $r_{01} = r_{10} = 1$, which gives no reflection of properties possessed by $X(t)$.

As mentioned earlier, Yong (1976) proved that $X(t)$ is ergodic if and only if $\overline{X}(n)$ is ergodic. In fact $X(t)$ and $\overline{X}(n)$ have the same limiting distribution π. That is

$$\lim_{n\to\infty} \overline{p}_{ij}(n) = \lim_{t\to\infty} p_{ij}(t) = \pi_j$$

independent of i for all j and $\sum_{j=0}^{\infty} \pi_j = 1$. We shall extend Yong's result to strong ergodicity.

To this end, we know that $P(t) = \exp Qt$. Since we are assuming $\sup_i |q_{ii}| = q<\infty$, for any $c>q$, $Q = c(\overline{P}-I)$. Hence

$$P(t) = \exp[ct(\overline{P}-I)] = [I + \sum_{k=1}^{\infty} \frac{(ct)^k}{k!} (\overline{P}-I)^k].$$

In general $\exp(A+B)$ need not equal $\exp(A)\cdot\exp(B)$ when A and B are matrices. Yet $\overline{P}\cdot I = I\cdot\overline{P}$, so that we will be able to rewrite $\exp[ct(\overline{P}-I)]$.

<u>Lemma 3.2:</u>

$$P(t) = \exp(-ct) \cdot \exp(ct\,\overline{P})$$

Proof:

$$P(t) = \exp(ct(\overline{P}-I)) = I + \sum_{k=1}^{\infty} \frac{(ct)^k}{k!} (\overline{P}-I)^k$$

$$= I + \sum_{k=1}^{\infty} \frac{(ct)^k}{k!} \sum_{\ell=0}^{k} \binom{k}{\ell} \overline{P}^{\ell} (-1)^{k-\ell} .$$

We will be assuming that the n-step transition probability matrix $\overline{P}^n$ may be written as $\overline{P}(n)$. Now we may write

$$P(t) = I + \sum_{k=1}^{\infty} \sum_{\ell=0}^{k} \frac{(ct)^{\ell}(-ct)^{k-\ell}}{(k-\ell)!\ell!} \cdot \overline{P}(\ell)$$

$$= I + \sum_{k=1}^{\infty} [\sum_{\ell=1}^{k} \frac{(ct)^{\ell}(-ct)^{k-\ell}\overline{P}(\ell)}{(k-\ell)!\ell!} + \frac{(-ct)^k}{k!} \cdot I]$$

$$= I + \sum_{\ell=1}^{\infty} [\sum_{k=\ell}^{\infty} \frac{(-ct)^{k-\ell}}{(k-\ell)!}] \frac{(ct)^{\ell}\overline{P}(\ell)}{\ell!} + \sum_{k=1}^{\infty} \frac{(-ct)^k}{k!} \cdot I$$

The interchange of the summations is valid since

$$|| \sum_{k=1}^{\infty} \sum_{\ell=0}^{\infty} \frac{(ct)^k(-1)^{k-\ell}\overline{P}(\ell)}{(k-\ell)!\ell!} || \leq \sum_{k=1}^{\infty} \sum_{\ell=0}^{\infty} \frac{(ct)^k}{\ell!(k-\ell)!}$$

$$= \sum_{k=1}^{\infty} (\sum_{\ell=0}^{k} \binom{k}{\ell}) \frac{(ct)^k}{k!}$$

$$\leq \exp(2ct).$$

That is to say the sum $\sum_{k=1}^{\infty} \sum_{\ell=0}^{\infty} \frac{(ct)^k(-1)^{k-\ell}\overline{P}(\ell)}{(k-\ell)!\ell!}$ is absolutely summable, so we may interchange the order of summation by Fubini's Theorem. Hence

$$P(t) = I + \sum_{\ell=1}^{\infty} \sum_{n=0}^{\infty} \frac{(-ct)^n}{n!} \frac{\overline{P}(\ell)(ct)^{\ell}}{\ell!} + (e^{-ct}-1)I$$

$$= e^{-ct} \cdot \sum_{\ell=1}^{\infty} \frac{\overline{P}(\ell)(ct)^{\ell}}{\ell!} + e^{-ct} I$$

$$= e^{-ct} \sum_{\ell=0}^{\infty} \frac{(ct)^{\ell}}{\ell!} \overline{P}(\ell)$$

$$= e^{-ct} \cdot \exp(ct\,\overline{P}).$$

Hence the proof is complete.

Since the limiting distributions for $X(t)$ and $\overline{X}(n)$ are equal, when either Markov chain is ergodic, we may define L to be a row constant stochastic matrix, each row of L being the common stationary distribution. Since we will be relating strong ergodicity for $X(t)$ in terms of strong ergodicity for $\overline{X}(n)$ we would like to relate $P(t)-L$ to $\overline{P}(n)-L$. Using Lemma 3.2 we can almost achieve a functional relationship.

Suppose we let $t=1$. By Lemma 3.2, $P(1)-L = e^{-c} \exp c\overline{P} - L$. Since both chains are ergodic we have

$$L^n = P(t)\cdot L = LP(t) = \overline{P}(m)\cdot L = L\overline{P}(m) = L$$

for all n, m, and t. Hence we may write

$$P(1)-L = e^{-c}\Big[\sum_{n=0}^{\infty} \frac{c^n\overline{P}(n)}{n!} - \sum_{n=1}^{\infty} \frac{c^n L^n}{n!} + \Big(\sum_{n=1}^{\infty} \frac{c^n}{n!}\Big)L\Big] - L$$

$$= e^{-c}\Big[I + \sum_{n=1}^{\infty} \frac{c^n}{n!} (\overline{P}(n)-L^n)\Big] + e^{-c}[e^c-1]L - L \qquad (3.1)$$

Since $\bar{P}(n) = \bar{P}^n$, and $\bar{P}$ and L commute we can easily show that

$$(\bar{P}(n)-L^n) = (\bar{P}-L)^n.$$

Replacing $\bar{P}(n)-L^n$ by $(\bar{P}-L)^n$ in (3.1) we have

$$P(1)-L = e^{-c}[\sum_{n=0}^{\infty} \frac{c^n(\bar{P}-L)^n}{n!}] - e^{-c}L$$

or

$$P(1)-L = e^{-c}\cdot\exp(c(\bar{P}-L)) - e^{-c}L . \tag{3.2}$$

The relationship given by (3.2) is not a strict functional relationship between P(1)-L and $\bar{P}$-L. Yet by using functional analytic techniques, we will be able to relate the spectral radius of $\bar{P}$-L to the spectral radius of P(1)-L. But first we need the following result.

Lemma 3.3:

If X(t) (or $\bar{X}(n)$) is ergodic, then

$$r(P(1)-L) \leq r(P(1)-L + e^{-c}L).$$

Proof:

By definition $r(P(1)-L) = \lim_{n\to\infty} ||(P(1)-L)^n||^{\frac{1}{n}}$. Consider

$$(P(1)-L)^{n+1} = (P(1)-L)^{n+1} + (P(1)-L)\cdot L\cdot e^{-nc}$$

$$= (P(1)-L)[(P(1)-L)^n + e^{-nc}\cdot L].$$

The first equality holds since $(P(1)-L)\cdot L = P(1)L-L^2 = L-L = \phi$. By induction on n we may easily show that

$(P(1)-L)^n + e^{-nc}L$ equals $(P(1)-L+e^{-c}L)^n$. Hence $(P(1)-L)^{n+1} = (P(1)-L)(P(1)-L+e^{-c}L)^n$. Therefore

$$||(P(1)-L)^n||^{\frac{1}{n}} = ||(P(1)-L)\cdot(P(1)-L+e^{-c}L)^{n-1}||^{\frac{1}{n}}$$

$$\leq ||P(1)-L||^{\frac{1}{n}} \cdot ||(P(1)-L+e^{-c}L)^{n-1}||^{\frac{1}{n-1}\cdot\frac{n-1}{n}}$$

$$\leq 2^{\frac{1}{n}} \cdot (||(P(1)-L+e^{-c}L)^{n-1}||^{\frac{1}{n-1}})^{\frac{n-1}{n}}$$

Hence $r(P(1)-L) = \lim\limits_{n\to\infty}||(P(1)-L)^n||^{\frac{1}{n}}$

$$\leq \lim_{n\to\infty}(||(P(1)-L+e^{-c}L)^{n-1}||^{\frac{1}{n-1}})^{\frac{n-1}{n}}.$$

Since $\frac{n-1}{n} \to 1$ as $n\to\infty$ we have that $r(P(1)-L) \leq r(P(1)-L+e^{-c}L)$. This completes the proof.

Rewriting the result given by (3.2) we have that

$$(P(1)-L) + e^{-c}L = e^{-c}\cdot\exp c(\overline{P}-L).$$

By Lemma 3.3 we have that

$$r(P(1)-L) \leq r(P(1)-L+e^{-c}L) = r(e^{-c}\cdot\exp c(\overline{P}-L)).$$

From Theorem 2.1 we know that $X(t)$ is strongly ergodic if and only if $r(P(1)-L)<1$. From the work of Isaacson and Luecke (1978) strong ergodicity for $\overline{X}(n)$ is equivalent to having $r(\overline{P}-L)<1$. Thus we shall show $r(\overline{P}-L)<1$ implies $r(P(1)-L)<1$, to obtain a characterization of strong ergodicity for $X(t)$. To accomplish this we need to show that

$r(\exp c(\bar{P}-L)) \leq \exp(c.r(\bar{P}-L))$.

To prove the above statement we consider the spectrum of $\exp(c(\bar{P}-L))$. By the Spectral Mapping Theorem we can show that

$$\sigma(\exp(c(\bar{P}-L))) = \exp(c.\sigma(\bar{P}-L)),$$

since $\exp(\cdot)$ is a continuous function (Rudin (1973)). We interpret $\exp(c.\sigma(\bar{P}-L))$ as $\{e^{c\lambda}: \lambda\varepsilon\sigma(\bar{P}-L)\}$. Hence

$$\begin{aligned} r(\exp(c(\bar{P}-L))) &= \sup\{|\lambda|: \lambda\varepsilon\sigma(\exp c(\bar{P}-L))\} \\ &= \sup\{|\lambda|: \lambda\varepsilon \exp(c.\sigma(\bar{P}-L))\} \\ &= \sup\{|e^{c\gamma}|: \gamma\varepsilon\sigma(\bar{P}-L)\} \\ &= \sup\{e^{cRe\gamma}: \gamma\varepsilon\sigma(\bar{P}-L)\}. \end{aligned}$$

The last equality is found by the relation, $|e^z| = e^{Rez}$ for any complex number z. Now for any $z\varepsilon\sigma(\bar{P}-L)$, $|z|^2 = (Rez)^2 + (Imz)^2$ hence $|Rez| \leq |z|$. Since $r(\exp c(\bar{P}-L)) = \sup\{e^{cRe\gamma}: \gamma\varepsilon\sigma(\bar{P}-L)\}$ we can see that $\sup\{Re\gamma: \gamma\varepsilon\sigma(\bar{P}-L)\} \leq \sup\{|Re\gamma|: \gamma\varepsilon\sigma(\bar{P}-L)\} \leq \sup\{|\gamma|: \gamma\varepsilon\sigma(\bar{P}-L)\} = r(\bar{P}-L)$. Therefore,

$$\begin{aligned} r(\exp(\bar{P}-L)) &= \sup\{e^{cRe\gamma}: \gamma\varepsilon\sigma(\bar{P}-L)\} \\ &= \exp(\sup\{c\cdot Re\gamma: \gamma\varepsilon\sigma(\bar{P}-L)\}) \\ &\leq \exp(cr(\bar{P}-L)). \end{aligned}$$

Therefore, we have the following result.

Lemma 3.4:

$$r(P(1)-L) \leq e^{-c} \exp(c\cdot r(\overline{P}-L)).$$

Proof:

From Lemma 3.3 $r(P(1)-L) \leq e^{-c}r(\exp c(\overline{P}-L))$. From the above argument $r(\exp c(\overline{P}-L) \leq \exp c(r(\overline{P}-L))$. Therefore $r(P(1)-L) \leq e^{-c}(\exp c\cdot r(\overline{P}-L))$, as desired.

Using the above lemma we can relate strong ergodicity for $X(t)$ in terms of strong ergodicity for $\overline{X}(n)$.

Theorem 3.1:

If $\overline{X}(n)$ is strongly ergodic then $X(t)$ is strongly ergodic.

Proof:

If $\overline{X}(n)$ is strongly ergodic then $r(\overline{P}-L) < 1$. By Lemma 3.4, $r(P(1)-L) \leq e^{-c} \exp(cr(\overline{P}-L))$. Since $-c(1-r(\overline{P}-L)) < 0$ we have that $r(P(1)-L) < 1$. Thus by Theorem 2.1, $X(t)$ is strongly ergodic.

In Chapter IV we shall prove the converse of Theorem 3.1. For now, though, we can prove the converse by adding assumptions. But first we need the following definition.

Definition 3.2:

Let A be an infinite dimensional matrix. If for some complex number λ there exists a sequence of unit vectors,

$\{x_n\}$, such that $\lim_{n\to\infty} ||(A-\lambda I)x_n|| = 0$, then λ is said to be an element of the approximate point spectrum of A. The set of all such λ's will be denoted by $\sigma_\pi(A)$.

It is well-known that the boundary of the spectrum is contained in the approximate point spectrum. Also the boundary of the spectrum contains all those points whose modulus is the spectral radius. Hence the approximate point spectrum contains this collection of points.

Applying these comments to our situation we have that, if there exists a γ, such that γ is real and $\gamma = r(\overline{P}-L)$ then it is easily shown that $r(\exp(c(\overline{P}-L))) = \exp(cr(\overline{P}-L))$. If we can show that $r(P(1)-L) = e^{-c}r(\exp c(\overline{P}-L))$ then the converse of Theorem 3.1 will hold. To this end, let $\lambda\varepsilon\sigma_\pi(P(1)-L + e^{-c}L)$. Hence there exists a sequence $\{x_n\}$ such that $||x_n|| = 1$ and $\lim_{n\to\infty}||(P(1)-L + e^{-c}L - \lambda I)x_n|| = 0$. Since

$$||L(P(1)-L+e^{-c}L-\lambda I)x_n|| \le ||L||\cdot||(P(1)-L+e^{-c}L-\lambda I)x_n||$$

we have that $\lim_{n\to\infty}||L(P(1)-L+e^{-c}L-\lambda I)x_n|| = 0$. Yet $L(P(1)-L) + e^{-c}L^2-\lambda L = (e^{-c}-\lambda)\cdot L$. Hence for $\lambda\neq e^{-c}$

$$\lim_{n\to\infty}||Lx_n|| = 0.$$

Thus $\lim_{n\to\infty}||(P(1)-L-\lambda I)x_n|| \le \lim_{n\to\infty}||(P(1)-L+e^{-c}L-\lambda I)x_n|| + \lim_{n\to\infty}||e^{-c}Lx_n|| = 0$. Thus by definition $\lambda\varepsilon\sigma_\pi(P(1)-L)$, for

$\lambda \neq e^{-c}$. Similarly we can show that for $0 \neq \lambda \varepsilon \sigma_{\pi}(P(1)-L)$ that $\lambda \varepsilon \sigma_{\pi}(P(1)-L+e^{-c}L)$. Hence the following set equivalence holds,

$$\{\lambda \colon \lambda \varepsilon \sigma_{\pi}(P(1)-L) \text{ or } \lambda = e^{-c}\} = \{\lambda \colon \lambda = 0$$

$$\text{or } \lambda \varepsilon \sigma_{\pi}(P(1)-L+e^{-c}L)\}.$$

From the comments after Definition 3.2 by taking the supremum we have

$$r(P(1)-L+e^{-c}L) = \max\{r(P(1)-L),\ e^{-c}\}.$$

Now let c be so large so that $e^{-c} < r(P(1)-L)$. And suppose that $r(\overline{P}-L)$ is an element of $\sigma_{\pi}(\overline{P}-L)$. Hence we have that

$$r(P(1)-L) = r(P(1)-L + e^{-c}L) = e^{-c}\exp(c \cdot r(\overline{P}-L)).$$

Hence the spectral radius of $P(1)-L$ is less than 1 if and only if the spectral radius of $\overline{P}-L$ is less than 1. Hence $X(t)$ is strongly ergodic if and only if $\overline{X}(n)$ is strongly ergodic.

In view of the above argument, one might conjecture, "the spectral radius of $\overline{P}-L$ is always an element of the spectrum of $\overline{P}-L$". By the following example the conjecture is false.

<u>Example 3.4:</u>

Suppose

$$Q = \begin{bmatrix} -1 & 1 & 0 \\ 0 & -1 & 1 \\ 1 & 0 & -1 \end{bmatrix}.$$

Let c=2 and then

$$\overline{P} = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}.$$

The powers of $\overline{P}$ generate a discrete time, irreducible, aperiodic, recurrent chain, $\overline{X}(n)$. Since $\delta(\overline{P}) = \frac{1}{2} \sup_{i,j} \sum_{n=0}^{2} |\overline{p}_{in} - \overline{p}_{jn}| = \frac{1}{2} \max\{1,1,1\} = \frac{1}{2}$, $\overline{X}(n)$ is strongly ergodic. The long run distribution is given by $\pi = (\frac{1}{3}, \frac{1}{3}, \frac{1}{3})$. With $L = \begin{bmatrix} \pi \\ \pi \\ \pi \end{bmatrix}$ we can easily show that $\sigma(\overline{P}-L) = \{0, \frac{1+i\sqrt{3}}{4}, \frac{1-i\sqrt{3}}{4}\}$. Thus $r(\overline{P}-L) = \frac{1}{2}$ and $\frac{1}{2}$ is not an element of the spectrum of $\overline{P}$-L. Hence the conditions to gain the converse of Theorem 3.1 are needed.

Note that when $r(P(1)-L) = e^{-c} \exp c \cdot r(\overline{P}-L)$, we have that the exponential rate of convergence of P(t) is a simple exponential function of the best possible geometric rate of convergence of $\overline{P}(n)$.

IV. STRONG ERGODICITY FOR CONTINUOUS TIME HOMOGENEOUS CHAINS USING MEAN VISIT TIMES

In Chapter III we showed that if the intensity of passage matrix, Q, satisfies certain regularity conditions then strong ergodicity for the continuous time chain, which Q determines, can be obtained from a discrete time Markov chain determined by Q. We also showed that strong ergodicity for the continuous time chain implies strong ergodicity for the discrete time chain when certain assumptions are met.

In this chapter we show that strong ergodicity for a continuous time, homogeneous, irreducible Markov chain is equivalent to strong ergodicity for the discrete time Markov chain without the added assumptions given in Chapter III.

Throughout this chapter let X(t) be a continuous time, homogeneous, irreducible Markov chain. Let the intensity of passage matrix, Q, satisfy $\sup_i\{|q_{ii}|\} = q<+\infty$. As in Chapter III we may write, for any c>q,

$$\overline{P} = I + \frac{Q}{c} \; .$$

The powers of $\overline{P}$, $\overline{P}^n = \overline{P}(n)$, generate a discrete time, homogeneous, aperiodic, irreducible discrete time Markov chain, $\overline{X}(n)$. Again if $\overline{X}(n)$ or X(t) is ergodic, then the other is ergodic with common limiting distribution π.

With the assumptions given above, the following results are consequences of the Renewal Theorem. For any $i \varepsilon S$

$$\lim_{n\to\infty} \bar{p}_{ji}(n) = \frac{1}{\sum_{k=1}^{\infty} k \cdot \bar{f}_{ii}(k)} \quad \text{for all } j\varepsilon S, \text{ if}$$

$$\sum_{k=1}^{\infty} k \cdot \bar{f}_{ii}(k) < +\infty$$

$$= 0 \text{ otherwise.}$$

Similarly for all $i \varepsilon S$

$$\lim_{t\to\infty} p_{ji}(t) = \frac{1}{-q_{ii}\int_0^{\infty} t dF_{ii}(t)} \quad \text{for all } j\varepsilon S \text{ if}$$

$$\int_0^{\infty} t \cdot dF_{ii}(t) < +\infty,$$

$$= 0 \text{ otherwise.}$$

Here $\bar{f}_{ii}(k)$ and $F_{ii}(t)$ are the first return probability to state i, for $\bar{X}(n)$, and the first return distribution to state i, for $X(t)$, as defined in Chapter I.

From the results given in Chapter I we know that

$$\bar{m}_{ii} = \sum_{k=1}^{\infty} k \cdot \bar{f}_{ii}(k) \quad \text{and} \quad m_{ii} = \int_0^{\infty} t dF_{ii}(t).$$

Hence, by the Renewal Theorem, for all $i \varepsilon S$

$$\lim_{n\to\infty} \bar{p}_{ji}(n) = \frac{1}{\bar{m}_{ii}} \quad \text{and} \quad \lim_{t\to\infty} p_{ji}(t) = \frac{1}{-q_{ii} \cdot m_{ii}}.$$

If we assume that $X(t)$ is ergodic then we know that $\lim_{n\to\infty} \bar{p}_{ji}(n) = \lim_{t\to\infty} p_{ji}(t) = \pi_i$ for all i independently of j with

$\sum_{j=0}^{\infty} \pi_j = 1$. Thus from the above result

$$\frac{1}{\bar{m}_{ii}} = \lim_{n\to\infty} \bar{p}_{ji}(n) = \pi_i = \lim_{t\to\infty} p_{ji}(t) = \frac{1}{-q_{ii}\cdot m_{ii}},$$

or that

$$m_{ii}\cdot(-q_{ii}) = \bar{m}_{ii}.$$ Hence we have the following.

Lemma 4.1:

$X(t)$ is positive recurrent if and only if $\bar{X}(n)$ is positive recurrent. In fact $m_{ii} = \frac{\bar{m}_{ii}}{-q_{ii}}$, for all i.

The importance of this lemma is realized from the fact that we may calculate the mean return time to state i from state i for the continuous time chain by considering only the discrete time chain.

Example 4.1:

Let $X(t)$ be a continuous time Markov chain with intensity matrix given by

$$Q = \begin{pmatrix} -1 & 0 & 1 & 0 \\ 2 & -2 & 0 & 0 \\ 0 & 0 & -4 & 4 \\ 0 & 6 & 0 & -6 \end{pmatrix}.$$

Let $c = 8$ and then

$$\overline{P} = \begin{bmatrix} \frac{7}{8} & 0 & \frac{1}{8} & 0 \\ \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 0 & \frac{3}{4} & 0 & \frac{1}{4} \end{bmatrix}.$$

The long run distribution for $\overline{P}$ is given by $\pi = (\frac{12}{23}, \frac{6}{23}, \frac{3}{23}, \frac{2}{23})$. Hence $\overline{m}_{11} = \frac{23}{12}$, $\overline{m}_{22} = \frac{23}{6}$, $\overline{m}_{33} = \frac{23}{3}$ and $\overline{m}_{44} = \frac{23}{2}$. From Lemma 4.1 we have $m_{11} = \frac{23}{12}$, $m_{22} = \frac{23}{12}$, $m_{33} = \frac{23}{12}$ and $m_{44} = \frac{23}{12}$.

In view of Lemma 4.1, one would hope that a simple functional relationship between m_{ii} and $\overline{m}_{ii}$ also holds for m_{ij} and $\overline{m}_{ij}$. As we shall show, this is the case, in fact $m_{ij} = \frac{\overline{m}_{ij}}{c}$ for all $i \neq j$. Yet, we will need some preliminary work.

For discrete time chains, mean visit times may be calculated by using generating functions (see Isaacson and Madsen (1976)). Since we will be dealing with both continuous and discrete time chains, we shall use a more general approach. That is, we will use the Laplace transforms of the functions $F_{ij}(t)$ and $p_{ij}(t)$, and of the sequences $\overline{p}_{ij}(n)$ and $\overline{f}_{ij}(n)$, to calculate mean visit times.

Definition 4.1:

a) $F_{ij}^*(u) = \int_0^\infty e^{-ut} dF_{ij}(t)$ = Laplace transform of $F_{ij}(t)$.

b) $P_{ij}^*(u) = \int_0^\infty e^{-ut} p_{ij}(t)\,dt$ = Laplace transform of $p_{ij}(t)$.

c) $\bar{f}_{ij}^*(u) = \sum_{k=0}^{\infty} e^{-uk} \bar{f}_{ij}(k)$ = Laplace transform of $\bar{f}_{ij}(k)$.

d) $\bar{p}_{ij}^*(u) = \sum_{k=0}^{\infty} e^{-uk} \bar{p}_{ij}(k)$ = Laplace transform of $\bar{p}_{ij}(n)$.

We now show that each of the transforms exists and is finite for each $u>0$. Since $|F_{ij}^*(u)| \leq \int_0^\infty |e^{-ut}| dF_{ij}(t) \leq \int_0^\infty dF_{ii}(t) = 1$, we have that $F_{ij}^*(u)$ exists and is finite for $u>0$. Similarly we can show that for $u>0$ $|P_{ij}^*(u)| \leq \frac{1}{u}$, $|\bar{f}_{ij}^*(u)| \leq 1$, and $\bar{p}_{ij}^*(u) \leq (1-e^{-u})^{-1}$.

The mean visit time from state i to state j, for $X(t)$, may be calculated by taking the derivative of $F_{ij}^*(u)$ at $u=0$. Similarly, the mean visit time from state i to state j, for $\bar{X}(n)$, may be calculated by finding $\lim_{u\to 0} \frac{d}{du} \bar{f}_{ij}^*(u)$. We shall show that $F_{ij}^*(u) = \bar{f}_{ij}^*(\ln(\frac{u+c}{c}))$, by relating $P_{ij}^*(u)$ to $\bar{p}_{ij}^*(u)$. Hence, we will be able to prove that $\bar{m}_{ij} = m_{ij}.c$.

The probability of visiting state j, from state i, at time n, using $\bar{X}(n)$, can be found by conditioning on the time of the first arrival to state j from state i. That is $\bar{p}_{ij}(n) = \sum_{k=1}^{n} \bar{f}_{ij}(k)\cdot\bar{p}_{jj}(n-k)$. This representation has the following continuous time analogue.

<u>Lemma 4.2</u>:

For all i and j

$$p_{ij}(t) = \delta_{ij}e^{q_{ii}.t} + \int_0^t p_{jj}(t-s)\,dF_{ij}(s).$$

<u>Proof</u>:

See Chung (1967).

We may now appeal to Laplace transforms to show that $F_{ij}^*(u)$ is a function of $P_{ij}^*(u)$ and $P_{ii}^*(u)$.

<u>Lemma 4.3</u>:

If $i \neq j$ then $P_{ij}^*(u) = P_{jj}^*(u)F_{ij}^*(u)$.

<u>Proof</u>:

By Lemma 4.2

$$p_{ij}(t) = \delta_{ij}.e^{q_{ii}.t} + \int_0^t p_{jj}(t-s)\,dF_{ij}(s)$$
$$= \int_0^t p_{jj}(t-s)\,dF_{ij}(s).$$

Taking the Laplace transform of both sides we find

$$P_{ij}^*(u) = \int_0^\infty e^{-ut} p_{ij}(t)dt = \int_0^\infty e^{-ut}[\int_0^t p_{jj}(t-s)dF_{ij}(s)]dt$$
$$= \int_0^\infty \int_0^t e^{-ut} p_{jj}(t-s)dF_{ij}(s)dt.$$

Since the terms of integration are nonnegative we may interchange the order of integration, by Tonelli's Theorem. Thus

$$P_{ij}^*(u) = \int_0^\infty [\int_s^\infty e^{-u(t-s)} p_{jj}(t-s)dt]e^{-us}dF_{ij}(s)$$

$$= \int_0^\infty [\int_0^\infty e^{-ux} p_{jj}(x)dx]e^{-us}dF_{ij}(s)$$

$$= P_{jj}^*(u) \cdot \int_0^\infty e^{-us}dF_{ij}(s)$$

$$= P_{jj}^*(u) \cdot F_{ij}^*(u).$$

Hence

$$F_{ij}^*(u) = \frac{P_{ij}^*(u)}{P_{jj}^*(u)} \text{ for all } i \neq j.$$

From the comment preceding Lemma 4.2 and by using the technique given in Lemma 4.3 we can show that

$$\bar{f}_{ij}^*(u) = \bar{p}_{ij}^*(u)/\bar{p}_{jj}^*(u).$$

Since $F_{ij}^*(u) = P_{ij}^*(u)/P_{jj}^*(u)$ and $\bar{f}_{ij}^*(u) = \bar{p}_{ij}^*(u)/\bar{p}_{jj}^*(u)$, if we can relate $P_{ij}^*(u)$ to $\bar{p}_{ij}^*(n)$ then we will be able to relate $F_{ij}^*(u)$ to $\bar{f}_{ij}^*(u)$. We relate $P_{ij}^*(u)$ to $\bar{p}_{ij}^*(u)$ by employing the relation $P(t) = e^{-ct} \exp(ct\,\bar{P})$ in the next lemma.

Lemma 4.4:

For all i and j

$$P_{ij}^{*}(u) = \frac{1}{u+c}\,\bar{p}_{ij}^{*}(\ln(\frac{u+c}{c})).$$

Proof:

$$P_{ij}^{*}(u) = \int_0^\infty e^{-ut} p_{ij}(t)\,dt$$

$$= \int_0^\infty e^{-ut}[e^{-ct} \cdot \sum_{n=0}^{\infty} \frac{(ct)^n}{n!}\,\bar{p}_{ij}(n)]\,dt$$

$$= \int_0^\infty \sum_{n=0}^{\infty} \frac{(ct)^n}{n!}\,\bar{p}_{ij}(n)\,e^{-(u+c)t}\,dt.$$

Again by Tonelli's Theorem we may interchange the order of integration and summation to find

$$P_{ij}^{*}(u) = \sum_{n=0}^{\infty} [\int_0^\infty t^n e^{-(u+c)t} dt]\,\frac{c^n}{n!}\,\bar{p}_{ij}(n)$$

$$= \sum_{n=0}^{\infty} [\int_0^\infty x^n e^{-x}\,dx]\,(\frac{c}{u+c})^n \cdot \frac{1}{u+c} \cdot \frac{1}{n!}\,\bar{p}_{ij}(n)$$

$$= \sum_{n=0}^{\infty} n!\,(\frac{c}{u+c})^n \cdot \frac{1}{u+c} \cdot \frac{1}{n!}\,\bar{p}_{ij}(n)$$

$$= \frac{1}{u+c} \sum_{n=0}^{\infty} \exp[-n \cdot \ln(\frac{u+c}{c})] \cdot \bar{p}_{ij}(n)$$

$$= \frac{1}{u+c} \cdot \bar{p}_{ij}^{*}(\ln(\frac{u+c}{c}))$$

From Lemma 4.3 $F_{ij}^*(u) = \dfrac{P_{ij}^*(u)}{P_{jj}^*(u)}$. From the previous lemma $P_{ij}^*(u) = \frac{1}{u+c} \cdot \bar{p}_{ij}^*(\ln(\frac{u+c}{c}))$ for all i and j. Hence by combining the last two results we find that

$$F_{ij}^*(n) = \frac{P_{ij}^*(u)}{P_{jj}^*(u)} = \frac{\frac{1}{u+c}\,\bar{p}_{ij}^*(\ln(\frac{u+c}{c}))}{\frac{1}{u+c}\,\bar{p}_{jj}^*(\ln(\frac{u+c}{c}))}$$

$$= \frac{\bar{p}_{ij}^*(\ln(\frac{u+c}{c}))}{\bar{p}_{jj}^*(\ln(\frac{u+c}{c}))} = \bar{f}_{ij}^*(\ln(\frac{u+c}{c})).$$

Thus the following is true.

<u>Lemma 4.5</u>:

If $i \neq j$ then $F_{ij}^*(u) = \bar{f}_{ij}^*(\ln(\frac{u+c}{c}))$.

Since we have a simple functional relationship between the Laplace transforms, we may differentiate to obtain the relationship between mean visit times. To this end, consider $\frac{d}{du} F_{ij}^*(u) = \frac{d}{du}\int_0^\infty e^{-ut} dF_{ij}(t) = \int_0^\infty (\frac{d}{du} e^{-ut}) dF_{ij}(t)$

$$= \int_0^\infty -te^{-ut} dF_{ij}(t)$$

Note that $\int_0^\infty |-t.e^{-ut}| dF_{ij}(t) < +\infty$ for each u so that the interchange of the order of differentiation and integration is valid, by the dominated convergence theorem.

For $i \neq j$ suppose that the mean visit time from state i to state j for X(t) is finite. That is

$$m_{ij} = \int_0^\infty t \cdot dF_{ij}(t) < +\infty.$$

Thus

$$\lim_{u\to 0} \frac{d}{du} F_{ij}^*(u) = \lim_{u\to 0} \int_0^\infty -te^{-ut} dF_{ij}(t) = -m_{ij}, \text{ by the}$$

dominated convergence theorem.

Now, since $F_{ij}^*(u) = \bar{f}_{ij}^*(\ln(\frac{u+c}{c}))$

$$\frac{d}{du}F_{ij}^*(u) = \frac{d}{du} \bar{f}_{ij}^*(\ln(\frac{u+c}{c}))$$

$$= \frac{d}{du}\left(\sum_{k=1}^{\infty} \exp(-k \ln(\frac{u+c}{c}))\cdot \bar{f}_{ij}(k)\right)$$

$$= -\frac{1}{u+c} \cdot \sum_{k=1}^{\infty} k\cdot (\frac{c}{u+c})^k \cdot \bar{f}_{ij}(k).$$

Since $\sum_{k=1}^{\infty} k(\frac{c}{u+c})^k < \infty$ for all $u>0$, the termwise differentiation of the series is justified. Since m_{ij} is finite and $\lim_{u\to 0^+} \frac{d}{du} F_{ij}^*(u) = -m_{ij}$, we have that

$$m_{ij} = \lim_{u\to 0^+} \frac{1}{u+c} \cdot \sum_{k=1}^{\infty} k\cdot (\frac{c}{u+c})^k \cdot \bar{f}_{ij}(k) < +\infty.$$

By Abel's Theorem we may interchange the order of the limiting operations. Hence $m_{ij} = \frac{1}{c} \cdot \sum_{k=1}^{\infty} k\cdot\bar{f}_{ij}(k)$. By definition $\sum_{k=1}^{\infty} k\cdot\bar{f}_{ij}(k) = \bar{m}_{ij}$, thus $\bar{m}_{ij} = c\cdot m_{ij}$. Hence we have that if the mean visit time from state i to state j for X(t) is finite then the mean visit time from i to j using $\bar{X}(n)$ is finite, in fact $\bar{m}_{ij} = c \cdot m_{ij}$.

Conversely, suppose $\bar{m}_{ij} = \sum_{k=1}^{\infty} k \cdot \bar{f}_{ij}(k) < +\infty$. Hence

$$\lim_{u \to 0^+} \sum_{k=0}^{\infty} k \cdot \exp(-k \ln(\frac{u+c}{c})) \cdot \bar{f}_{ij}(k) = \bar{m}_{ij}$$

$$= c \cdot \lim_{u \to 0^+} \int_0^{\infty} te^{-ut} dF_{ij}(t).$$

Since $\lim_{u \to 0^+} \int_0^{\infty} t \cdot e^{-ut} dF_{ij}(t) < +\infty$, by Fatou's Lemma

$$c \cdot m_{ij} = c \cdot \int_0^{\infty} tdF_{ij}(t) \leq c \cdot \lim_{u \to 0^+} \int_0^{\infty} e^{-ut} \cdot tdF_{ij}(t) = \bar{m}_{ij}.$$

Hence if $\bar{m}_{ij} < +\infty$ then $m_{ij} < +\infty$. Hence by the first part we may write $m_{ij} = \frac{\bar{m}_{ij}}{c}$. Thus we have the following theorem.

<u>Theorem 4.1</u>:

The mean visit time from i to j, $i \neq j$, using $X(t)$ is finite if and only if the mean visit time from i to j for $\bar{X}(n)$ is finite. In this situation $m_{ij} = \frac{\bar{m}_{ij}}{c}$.

At the outset of this chapter we said that we would show the equivalence of strong ergodicity between $X(t)$ and $\bar{X}(n)$. Using Theorem 4.1 and Lemma 4.1 this is easily found. Because strong ergodicity for a Markov chain, discrete or continuous, is equivalent to having the supremum of the mean visit times, to some positive recurrent state, bounded over the starting states (Issacson and Arnold (1978); Huang and Isaacson (1976)).

Theorem 4.2:

If $X(t)$ has at least one positive recurrent state, then $X(t)$ is strongly ergodic if and only if $\overline{X}(n)$ is strongly ergodic.

Proof:

If state j is positive recurrent for $X(t)$, then by Lemma 4.1 state j is positive recurrent for $\overline{X}(n)$. Since the chain, $X(t)$, is irreducible all states of $X(t)$ and $\overline{X}(n)$ are positive recurrent. Hence $m_{ij} < \infty$ and $\overline{m}_{ij} < \infty$ for all i and j. Suppose $X(t)$ is strongly ergodic. From the comment preceding this theorem, this is equivalent to having

$$\sup_i m_{ij} < \infty.$$

From Lemma 4.2 and Theorem 4.1, $m_{ii} = \frac{\overline{m}_{ii}}{-q_{ii}}$ and $m_{ij} = \frac{\overline{m}_{ij}}{c}$ for $i \neq j$. Hence $\sup_i m_{ij} = \max\{\sup_{i \neq j} \frac{\overline{m}_{ij}}{c}, \frac{-\overline{m}_{jj}}{q_{jj}}\}$. Hence

$$\sup_i \overline{m}_{ij} < +\infty.$$

Since $\sup_i \overline{m}_{ij} < +\infty$, $\overline{X}(n)$ is strongly ergodic.

The converse is proved similarly.

In Chapter III we were able to show that if $\overline{X}(n)$ is strongly ergodic then $X(t)$ is strongly ergodic. This was accomplished by showing $r(\overline{P}-L) < 1$ implies $r(P(1)-L) < 1$.

In light of Theorem 4.2 we have the following.

Corollary 4.1:

The spectral radius of $P(1)-L$ is less than 1 if and only if the spectral radius of $\overline{P}-L$ is less than 1.

Proof:

Strong ergodicity for $X(t)$ is equivalent to having $r(P(1)-L) < 1$.

V. RATES OF CONVERGENCE FOR CONTINUOUS TIME HOMOGENEOUS CHAINS

Pitman (1974) used mean visit times to find new results on the uniform rate of convergence of discrete time Markov chains. He made no mention of the extension of his results to the case of continuous time. By appealing to the functional relationship between $P(t)$ and $\overline{P}$, given in Chapter III, and the functional relationship between the mean visit times for $X(t)$ and $\overline{X}(n)$ we can extend his results to the continuous time framework.

We shall assume throughout this chapter that $X(t)$ will be an irreducible, positive recurrent, homogeneous continuous time Markov chain on the state space $S = \{0,1,2,\ldots\}$. $P(t)$ will be the transition probability matrix and Q will be the intensity of passage matrix. Also, as in Chapter III, assume

$$\sup_i\{|q_{ii}|\} = q < +\infty,$$

and hence we may write, for $c>q$,

$$\overline{P} = I + \frac{Q}{c}.$$

Again, $\overline{P}$ will generate a discrete time, homogeneous, aperiodic, positive recurrent Markov chain, $\overline{X}(n)$. In fact we may write

$$P(t) = [I + \sum_{n=1}^{\infty} \frac{(ct)^n}{n!} \bar{P}(n)]e^{-ct},$$

where we are assuming that $\bar{P}(n)$ is $\bar{P}^n$.

Having $X(t)$ irreducible and positive recurrent implies the chain is ergodic. Hence $\bar{X}(n)$ is ergodic, also $X(t)$ and $\bar{X}(n)$ have a common limiting distribution, $\pi = (\pi_0,\pi_1, \ldots)$. Thus

$$\lim_{t\to\infty} p_{ij}(t) = \pi_j = \lim_{n\to\infty} \bar{p}_{ij}(n)$$

for all j independent of i. We formulate this fact in terms of initial distributions.

<u>Definition 5.1</u>:

A vector $\lambda = (\lambda_0,\lambda_1,\ldots)$ is said to be a probability distribution on $S = \{0,1,2,\ldots\}$ if $\lambda_j \geq 0$ for all j and $\sum_{j=0}^{\infty} \lambda_j = 1$.

Given a probability distribution λ on S, we may think of λ as an initial distribution for $X(t)$ or $\bar{X}(t)$. That is, set

$$p_{\lambda j}(t) = (\lambda P(t))_j = \sum_{i=0}^{\infty} \lambda_i p_{ij}(t)$$

and

$$\bar{p}_{\lambda j}(n) = (\lambda \bar{P}(n))_j = \sum_{i=0}^{\infty} \lambda_i \bar{p}_{ij}(n).$$

Thus $p_{\lambda j}(t)$ $(\bar{p}_{\lambda j}(n))$ is the probability that at time $t(n)$ the chain $X(t)$ ($\overline{X}(n)$) occupies state j, given that at time 0 the probability of occupying state i is λ_i. Hence

$$p_{\lambda j}(t) = P_\lambda(X(t)=j)$$

and

$$\overline{p}_{\lambda j}(n) = P_\lambda(\overline{X}(n)=j).$$

Therefore $\lambda P(t)$ ($\lambda\overline{P}(n)$) is the distribution of $X(t)$ ($\overline{X}(n)$), given λ.

The concept of initial distributions can be tied in with ergodicity. Let λ be an initial distribution for $\overline{X}(n)$. Consider $||\lambda\overline{P}(n)-\pi||$ where π is the long run distribution for $\overline{X}(n)$. Now

$$||\lambda\overline{P}(n)-\pi|| = \sum_{j=0}^{\infty} |\overline{p}_{\lambda j}(n)-\pi_j|$$

$$= \sum_{j=0}^{\infty} \Big| \sum_{i=0}^{\infty} \lambda_i \overline{p}_{ij}(n)-\pi_j \Big|$$

$$= \sum_{j=0}^{\infty} \Big| \sum_{i=0}^{\infty} \lambda_i (\overline{p}_{ij}(n)-\pi_j) \Big|$$

$$\leq \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \lambda_i |\overline{p}_{ij}(n)-\pi_j|$$

$$= \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} |\overline{p}_{ij}(n)-\pi_j| \lambda_i$$

$$\leq \sum_{i=0}^{\infty} \lambda_i \Big(\sum_{j=0}^{\infty} \overline{p}_{ij}(n) + \sum_{j=0}^{\infty} \pi_j \Big)$$

$$\leq 2 \cdot \sum_{i=0}^{\infty} \lambda_i = 2.$$

From the dominated convergence theorem we then have

$$\lim_{n\to\infty}||\lambda\overline{P}(n)-\pi|| = \sum_{j=0}^{\infty}\sum_{i=0}^{\infty}\lambda_i(\lim_{n\to\infty}\overline{p}_{ij}(n)-\pi_j) = 0.$$

Thus for any initial distribution on $\overline{X}(n)$, the distribution of $\overline{X}(n)$ converges to the limiting distribution for $\overline{X}(n)$. Note that, by the triangle inequality for $||\ ||$, for any two initial distributions λ and μ

$$\lim_{n\to\infty}||\lambda\overline{P}(n)-\mu\overline{P}(n)|| = 0.$$

Naturally similar amplifications can be made concerning $X(t)$. That is, for any two initial distributions λ and μ

$$\lim_{t\to\infty}||\lambda P(t)-\mu P(t)|| = 0.$$

The results in the previous paragraph are immediate consequences of ergodicity. Two of Pitman's results, to be stated next, strengthen the result given above for the discrete time chain $\overline{X}(n)$. We state Pitman's results in terms of $\overline{X}(n)$, yet the theorem holds for any discrete time chain possessing the same properties as $\overline{X}(n)$.

<u>Theorem 5.1</u>: (Pitman)

Suppose $\overline{X}(n)$ is an irreducible, aperiodic, positive recurrent, discrete time, homogeneous Markov chain. Let λ and μ be any two initial distributions on the state space $S = \{0,1,2,...\}$. Suppose λ and μ are such that for all j

$$\bar{m}_{\lambda j} = \sum_{i=0}^{\infty} \lambda_i \bar{m}_{ij} \text{ and } \bar{m}_{\mu j} = \sum_{i=0}^{\infty} \mu_i \bar{m}_{ij}$$

are finite. Then

$$\lim_{n\to\infty} n\cdot||\lambda\bar{P}(n)-\mu\bar{P}(n)|| = 0 \tag{5.1}$$

and

$$\sum_{n=1}^{\infty} ||\lambda\bar{P}(n)-\mu\bar{P}(n)|| < \infty \,. \tag{5.2}$$

By assuming the added conditions that $\bar{m}_{\lambda j}$ and $\bar{m}_{\mu j}$ are finite, Pitman's work gives stronger results concerning the convergence of $||\lambda\bar{P}(n)-\mu\bar{P}(n)||$. Statement (5.2) implies that $||\lambda\bar{P}(n)-\mu\bar{P}(n)||\to 0$ as $n\to\infty$, and statement (5.1) says that a common rate of convergence can be taken to be $\frac{1}{n}$.

Having $\bar{m}_{\lambda j}<\infty$ may be interpreted as having the "average mean waiting time" to visit state j finite. By assumption $\bar{m}_{ij}<+\infty$, yet we may not have, in some cases, $\bar{m}_{\lambda j}<+\infty$. As Pitman observed, though, if $\bar{m}_{\lambda j}$ is finite for some j then $\bar{m}_{\lambda j}$ is finite for all j.

As indicated at the start of this chapter, Pitman's results will be extended to the continuous time situation.

<u>Theorem 5.2</u>:

Let X(t) be a continuous time, homogeneous, irreducible, positive recurrent Markov chain. Let λ and μ be two initial distributions on $S = \{0,1,2,...\}$ such that for all j $m_{\lambda j}$

and $m_{\mu j}$ are finite. Then

$$\lim_{t\to\infty} t\cdot||\lambda P(t)-\mu P(t)|| = 0 \tag{5.3}$$

and

$$\int_0^\infty ||\lambda P(t)-\mu P(t)||\cdot dt < \infty \tag{5.4}$$

<u>Proof</u>:

To show (5.3) and (5.4) we appeal to the relationship between X(t) and $\overline{X}(n)$. That is, since X(t) is irreducible, and positive recurrent we have that $\overline{X}(n)$ is irreducible, aperiodic, and positive recurrent. Since we assumed that λ and μ are such that $m_{\lambda j}$ and $m_{\mu j}$ are finite for all j, we shall show that $\overline{m}_{\lambda j}$ and $\overline{m}_{\mu j}$ are finite. From the work done in Chapter IV we have

$$m_{ij} = \frac{\overline{m}_{ij}}{c} \text{ for all } i\neq j \quad \text{and} \quad m_{ii} = \frac{\overline{m}_{ii}}{-q_{ii}} .$$

Hence

$$\overline{m}_{\lambda j} = \sum_{i=0}^{\infty} \lambda_i \overline{m}_{ij} = \sum_{i\neq j} \lambda_i \overline{m}_{ij} + \lambda_j \overline{m}_{jj} .$$

$$= \sum_{i\neq j} \lambda_i c\cdot m_{ij} + \lambda_j (-q_{jj})\cdot m_{jj}$$

$$= c \sum_{i=0}^{\infty} \lambda_i m_{ij} + (-\lambda_j)(q_{jj}+c) m_{jj}$$

$$= cm_{\lambda j} - \lambda_j m_{jj} (q_{jj}+c) < \infty.$$

Thus $\overline{m}_{\lambda j}$ and $\overline{m}_{\mu j}$ are finite for all j. Hence the conditions of Theorem 5.1 are satisfied. Thus

$$\lim_{n\to\infty} ||\lambda\overline{P}(n)-\mu\overline{P}(n)||\cdot n = 0$$

and

$$\sum_{n=1}^{\infty} ||\lambda\overline{P}(n)-\mu\overline{P}(n)||<\infty.$$

Now consider $t||\lambda P(t)-\mu P(t)|| = t||(\lambda-\mu)P(t)||$

$$= t\cdot e^{-ct}\cdot||(\lambda-\mu)\cdot \sum_{n=0}^{\infty} \frac{(ct)^n}{n!} \overline{P}(n)||$$

$$\leqq t\cdot e^{-ct} \sum_{n=0}^{\infty} \frac{(ct)^n}{n!} ||\lambda\overline{P}(n)-\mu\overline{P}(n)||$$

$$= \frac{e^{-ct}}{c} \cdot \sum_{n=0}^{\infty} \frac{(ct)^{n+1}}{(n+1)!} (n+1)\cdot||\lambda\overline{P}(n)-\mu\overline{P}(n)||$$

$$= \frac{e^{-ct}}{c} \sum_{n=1}^{\infty} \frac{(ct)^n}{n!} n\cdot||\lambda\overline{P}(n-1)-\mu\overline{P}(n-1)||$$

Since $n\cdot||\lambda\overline{P}(n)-\mu\overline{P}(n)|| \to 0$ and $||\lambda\overline{P}(n)-\mu\overline{P}(n)|| \to 0$, for all $\varepsilon>0$ there exists an $N = N(\varepsilon)$ such that for $n \geqslant N$

$$(n-1)\cdot||\lambda\overline{P}(n-1)-\mu\overline{P}(n-1)|| + ||\lambda\overline{P}(n-1)-\mu\overline{P}(n-1)|| < \frac{\varepsilon\cdot c}{2} .$$

Thus $n||\lambda\overline{P}(n-1)-\mu\overline{P}(n-1)|| < \frac{\varepsilon\cdot c}{2}$. Now

$$t\cdot||\lambda P(t)-\mu P(t)|| \leq \frac{e^{-ct}}{c}[\sum_{n=1}^{N-1} \frac{(ct)^n}{n!}||\lambda\overline{P}(n)-\mu\overline{P}(n)||$$

$$+ \sum_{n=N}^{\infty} \frac{(ct)^n}{n!} \cdot \frac{\varepsilon\cdot c}{2}]$$

$$\leq \frac{e^{-ct}}{c} \sum_{n=1}^{N-1} \frac{(ct)^n}{n!}\cdot 2 + \frac{\varepsilon}{2}$$

$$< \frac{2e^{-ct}}{c} (N-1)\cdot(ct)^N + \frac{\varepsilon}{2}.$$

Since $e^{-ct}(ct)^N \to 0$ as $t \to \infty$, there exists a $T = T(\varepsilon)$ such that for $t > T$

$$\frac{2e^{-ct}(N-1)(ct)^N}{c} < \frac{\varepsilon}{2}.$$

Hence for any $\varepsilon > 0$ there exists a T such that for $t > T$

$$t \cdot ||\lambda P(t) - \mu P(t)|| < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus $\lim_{t\to\infty} t||\lambda P(t) - \mu P(t)|| = 0$. Hence statement (5.3) holds.

To prove (5.4) we appeal to the same techniques used inthe proof of (5.3). To this end

$$||\lambda P(t) - \mu P(t)|| \leq e^{-ct} \cdot \sum_{n=0}^{\infty} \frac{(ct)^n}{n!} ||\lambda \overline{P}(n) - \mu \overline{P}(n)||.$$

Integrating both sides we find

$$\int_0^\infty ||\lambda P(t) - \mu P(t)|| dt \leq \int_0^\infty e^{-ct} \sum_{n=0}^{\infty} \frac{(ct)^n}{n!} ||\lambda \overline{P}(n) - \mu \overline{P}(n)|| dt$$

Since each term of the sum is nonnegative we may interchange the order of integration and summation. So

$$\int_0^\infty ||\lambda P(t) - \mu P(t)|| dt \leq \sum_{n=0}^{\infty} ||(\lambda - \mu)\overline{P}(n)|| \int_0^\infty \frac{e^{-ct}(ct)^n}{n!} dt.$$

The function $\frac{c \cdot e^{-ct}(ct)^n}{n!}$ is the density of a gamma random variable with parameters n and c. Hence

$$\int_0^\infty \frac{e^{-ct}(ct)^n}{n!} dt = \frac{1}{c} \quad \text{for } n = 0,1,2,\ldots .$$

thus

$$\int_0^\infty ||\lambda P(t)-\mu P(t)||dt \leq \frac{1}{c}\cdot \sum_{n=0}^{\infty} ||\lambda\overline{P}(n)-\mu\overline{P}(n)||$$

$$= \frac{1}{c} \cdot ||\lambda-\mu||+\frac{1}{c}\cdot \sum_{n=1}^{\infty} ||\lambda\overline{P}(n)-\mu\overline{P}(n)||$$

$$< \frac{2}{c}+\frac{1}{c}\cdot \sum_{n=1}^{\infty} ||\lambda\overline{P}(n)-\mu\overline{P}(n)|| < \infty.$$

Thus the proof of the theorem is complete.

From Theorem 5.1 we know that $\lambda \sum_{k=1}^{n} \overline{P}(k)-\mu \sum_{k=1}^{n} \overline{P}(k)$ converges. For any initial distribution, $\lambda\cdot \sum_{k=1}^{n} \overline{P}(k)$ represents the "expected occupation time" measure for the chain $\overline{X}(n)$. To see this, consider for any state j the random variable

$$\overline{Z}_{\lambda j}(n) = \begin{cases} 1 & \text{if } X(n) = j, \text{ given } \lambda \\ 0 & \text{otherwise} \end{cases}$$

The expression $\sum_{k=1}^{n} \overline{Z}_{\lambda j}(n)$ represents the random variable which counts the number of visits to state j in the first n steps of $\overline{X}(n)$ given λ. Since $E(\overline{Z}_{\lambda j}(n)) = \overline{p}_{\lambda j}(n)$, $\sum_{k=1}^{n} \overline{p}_{\lambda j}(k)$ is the expected number of visits to state j in the first n steps. Hence the difference

$$\lambda \sum_{k=1}^{n} \overline{P}(k)-\mu \sum_{k=1}^{n} \overline{P}(k)$$

represents the difference of the expected number of visits for each state given λ and the expected number given μ.

Under the same conditions as Theorem 5.1, Pitman gave an explicit form for $\lim_{n\to\infty} \sum_{k=1}^{n} (\lambda\overline{P}(k)-\mu\overline{P}(k))$. Before giving Pitman's result, we need another definition.

Definition 5.2:

An additive set function ν defined on the power set of $S' = \{0,1,2,\ldots\}$ will be called a signed measure on S. Let $\nu_j = \nu(\{j\})$ on singleton sets $\{j\} \subseteq S$.

We now give Pitman's expression for $\lim_{n\to\infty} \sum_{k=1}^{n} (\lambda\overline{P}(k)-\mu\overline{P}(k))$.

Theorem 5.3: (Pitman)

Under the same conditions as Theorem 5.1 there exists a signed measure $\overline{\nu}$ on S such that $\overline{\nu}(S) = 0$ and

$$\lim_{n\to\infty} \left|\left| \sum_{k=1}^{n} \lambda\overline{P}(k) - \sum_{k=1}^{n} \mu\overline{P}(k)] - \overline{\nu}\right|\right| = 0.$$

Also ν is given by $\nu_j = \dfrac{\overline{m}_{\mu j}-\overline{m}_{\lambda j}}{\overline{m}_{jj}}$.

Elaborating on the form of ν, we have if $\overline{m}_{\lambda j}$ and $\overline{m}_{\mu j}$ are both finite then

$$\sum_{n=1}^{\infty} (\overline{p}_{\lambda j}(n)-\overline{p}_{\mu j}(n)) = \frac{\overline{m}_{\mu j}-\overline{m}_{\lambda j}}{\overline{m}_{jj}} .$$

Hence the difference between the expected number of visits to state j, given λ and μ, in n steps converges to $\dfrac{\overline{m}_{\mu j}-\overline{m}_{\lambda j}}{\overline{m}_{jj}}$.

We shall extend Theorem 5.3 for continuous time chains.

But first we need to formulate the analogous expression for the "expected occupation time measure".

It will be understood that

$$\int_0^t \lambda P(s)ds = \left(\int_0^t p_{\lambda 0}(s)ds, \int_0^t p_{\lambda 1}(s)ds, \ldots\right) .$$

We may interpret $\int_0^t p_{\lambda j}(s)ds$ in the same manner as $\sum_{k=1}^{n} \bar{p}_{\lambda j}(k)$. That is define

$$Z_{\lambda j}(s) = \begin{cases} 1 & \text{if } X(s) = j \text{ given } \lambda \\ 0 & \text{otherwise.} \end{cases}$$

Since $E(Z_{\lambda j}(s)) = p_{\lambda j}(s)$ we may view $\int_0^t p_{\lambda j}(s)ds$ as the expected amount of time the chain $X(t)$ spends in state j in t units of time, given λ. Thus $\int_0^t (p_{\lambda j}(s)-p_{\mu j}(s))ds$ is the difference in the expected amount of time spent in state j given λ and the expected amount of time given μ.

We now give a closed form for

$$\lim_{t\to\infty} \int_0^t \lambda P(s)ds - \int_0^t \mu P(s)ds.$$

Theorem 5.4:

Under the same conditions as Theorem 5.2, there exists a signed measure ν on S, with $\nu(S) = 0$, such that

$$\lim_{t\to\infty} \left|\left|\int_0^t \lambda P(s)ds - \int_0^t \mu P(s)ds - \nu\right|\right| = 0.$$

Also ν is given by $\nu_j = \dfrac{m_{\mu j}-m_{\lambda j}-(\mu_j-\lambda_j)m_{jj}}{-q_{jj}m_{jj}}$

Proof:

As in the proof of Theorem 5.2, the assumptions placed on $X(t)$ imply the conditions of Theorem 5.3. Hence for $\overline{X}(n)$, there exists a signed measure $\overline{\nu}$, with $\overline{\nu}_j = \frac{\overline{m}_{\mu j}-\overline{m}_{\lambda j}}{\overline{m}_{jj}}$, such that

$$\lim_{n\to\infty}||(\sum_{k=1}^{n} \lambda\overline{P}(k) - \sum_{k=1}^{n} \mu\overline{P}(k)) - \overline{\nu}|| = 0.$$

Consider

$$||\int_0^t \lambda P(s)\,ds - \int_0^t \nu P(s)\,ds|| = ||\int_0^t (\lambda-\mu)P(s)\,ds||$$

$$= ||\int_0^t e^{-cs} \sum_{k=0}^{\infty} \frac{(cs)^k}{k!}(\lambda\overline{P}(k)-\mu\overline{P}(k))\,ds||$$

From Theorem 5.2 we know the above expression in the norm is absolutely convergent we may interchange the order of integration and summation. Hence

$$||\int_0^t \lambda P(s)\,ds - \int_0^t \mu P(s)\,ds|| = \frac{1}{c}||\sum_{k=0}^{\infty} [\lambda\overline{P}(k)-\mu\overline{P}(k)]$$

$$\cdot \int_0^t e^{-cs} \frac{(cs)^k}{k!}\, c\,ds||$$

$$= \frac{1}{c}||\cdot\sum_{k=0}^{\infty} (\lambda\overline{P}(k)-\mu\overline{P}(k))P(X_{k,c}\leq t)||.$$

Here $X_{k,c}$ is a gamma random variable with parameters k and c. Now $||\sum_{k=0}^{\infty} (\lambda\overline{P}(k)-\mu\overline{P}(k))P(X_{k,c}\leq t)|| \leq \sum_{k=0}^{\infty} ||\lambda\overline{P}(k)-\mu\overline{P}(k)|| < +\infty$.

Hence the dominated convergence theorem applies, so

$$\lim_{t\to\infty} \sum_{k=0}^{\infty} P(X_{k,c}\le t)(\lambda\overline{P}(k)-\mu\overline{P}(k)) = \sum_{k=0}^{\infty}(\lambda\overline{P}(k)-\mu\overline{P}(k)).$$

We know that $\sum_{k=1}^{\infty}(\lambda\overline{P}(k)-\mu\overline{P}(k)) = \overline{\nu}$. Define $\overline{\nu}' = \overline{\nu}+(\lambda-\mu)$ and hence $\sum_{k=0}^{\infty}(\lambda\overline{P}(k)-\mu\overline{P}(k)) = \overline{\nu}+(\lambda-\mu) = \overline{\nu}'$. Define $\nu = \frac{\overline{\nu}'}{c}$ and consider

$$\lim_{t\to\infty}\left|\left|\lambda\int_0^t P(s)ds-\mu\int_0^t P(s)ds-\nu\right|\right|$$

$$= \lim_{t\to\infty}\frac{1}{c}\cdot\left|\left|\sum_{k=0}^{\infty}P(X_{k,c}\le t)(\lambda\overline{P}(k)-\mu\overline{P}(k))-\overline{\nu}'\right|\right|$$

$$= \frac{1}{c}\left|\left|\sum_{k=0}^{\infty}(\lambda\overline{P}(k)-\mu\overline{P}(k))-\overline{\nu}'\right|\right|$$

$$= \frac{1}{c}||\nu'-\nu'|| = 0.$$

Hence $\int_0^t \lambda P(s)ds-\int_0^t \mu P(s)ds$ converges to ν. All that remains to be shown is that $\nu(S) = 0$ and to give the form of ν. To this end

$$\nu(S) = \sum_{j=0}^{\infty}\nu_j = \frac{1}{c}\cdot\sum_{j=0}^{\infty}\overline{\nu}_j' = \frac{1}{c}\sum_{j=0}^{\infty}(\overline{\nu}_j+\lambda_j-\mu_j)$$

$$= \frac{1}{c}\left(\sum_{j=0}^{\infty}\overline{\nu}_j + \sum_{j=0}^{\infty}\lambda_j - \sum_{j=0}^{\infty}\mu_j\right)$$

$$= \frac{1}{c}(0+1-1) = 0.$$

Now $\nu_j = \frac{1}{c} \cdot \bar{\nu}_j' = \frac{1}{c}(\bar{\nu}_j+\lambda_j-\mu_j)$

$$= \frac{1}{c} \cdot \frac{\bar{m}_{\mu j}-\bar{m}_{\lambda j}}{\bar{m}_{jj}} + \frac{1}{c}(\lambda_j-\mu_j)$$

$$= \frac{1}{c\bar{m}_{jj}} \sum_{i=0}^{\infty} (\mu_i-\lambda_i)\bar{m}_{ij} + \frac{1}{c}(\lambda_j-\mu_j)$$

$$= \frac{-1}{c\cdot m_{jj}q_{jj}}\Big[\sum_{i\neq j} (\mu_i-\lambda_i)m_{ij}\cdot c + (\mu_j-\lambda_j)m_{jj}(-q_{jj})\Big]$$

$$+ \frac{1}{c}(\lambda_j-\mu_j)$$

$$= \frac{\sum_{i\neq j} (\mu_i-\lambda_i)m_{ij}}{-q_{jj}m_{jj}}$$

$$= \frac{m_{\mu j}-m_{\lambda j}-(\mu_j-\lambda_j)m_{jj}}{-q_{jj}m_{jj}}$$

VI. STRONG ERGODICITY FOR CONTINUOUS TIME NONHOMOGENEOUS CHAINS USING MEAN VISIT TIMES

From the work of Huang and Isaacson (1976), and Isaacson and Arnold (1978) we know that mean visit times play an important role in the characterization of strong ergodicity for discrete time nonhomogeneous and continuous time homogeneous Markov chains. The purpose of this section is to prove that for a continuous time nonhomogeneous Markov chain uniform strong ergodicity may be characterized by properties held by the mean visit times.

For this chapter $X(t)$ will denote a nonhomogeneous continuous time Markov chain on the state space $S = \{0,1,2,...\}$. $P(s,t)$ will represent the matrix of transition probability functions. The elements of $P(s,t)$ are given by

$$(P(s,t))_{ij} = p_{ij}(s,t) = P(X(t)=j|X(s)=i)$$

for all i and j in S. We also assume that for all $s\leq t$, and every i in S, $\sum_{j=0}^{\infty} p_{ij}(s,t)=1$.

The transition functions satisfy the Chapman Kolmogorov equations,

$$p_{ij}(s,t) = \sum_{k=0}^{\infty} p_{ik}(s,u)p_{kj}(u,t)$$

which hold for all $s<u<t$. This reflects the fact that a transition from state i, starting at time s, to state j,

at time t, must pass through some intermediate state k, at time u.

As in the case of continuous time homogeneous Markov chains, we may define, for each starting time, the transition intensity. The transition intensities will have the same probabilistic interpretation, but will depend on the starting time. To this end, define

$$q_{ii}(s) = \lim_{h\to 0^+} \frac{p_{ii}(s,s+h)-1}{h}$$

as the intensity of passage out of state i at time s. Define

$$q_{ij}(s) = \lim_{h\to 0^+} \frac{p_{ij}(s,s+h)}{h}$$

as the intensity of passage from state i to state j starting at time s. We shall assume that the family of functions $\{q_{ij}(s)\}$, is continuous. Also for fixed s and i $\sum_{j=0}^{\infty} q_{ij}(s)=0$.

If for fixed j, the passage to the limit in $p_{ij}(s,s+h)$ is uniform with respect to i, we obtain the well-known Kolmogorov differential equations. That is to say,

$$\frac{\partial}{\partial t} p_{ij}(s,t) = \sum_{k=0}^{\infty} p_{ik}(s,t)q_{kj}(t) \tag{6.1}$$

$$\frac{\partial}{\partial s} p_{ij}(s,t) = -\sum_{k=0}^{\infty} q_{ik}(s)\cdot p_{kj}(s,t). \tag{6.2}$$

We shall assume for all i and s

$$0 < b \le |q_{ii}(s)| \le B < \infty$$

This assumption is required for the following reasons. First we may write the integrated form of (6.1) as

$$p_{ij}(s,t) = \delta_{ij} + \int_s^t \left(\sum_{k=0}^{\infty} p_{ik}(s,u) \cdot q_{kj}(u) \right) du \qquad (6.3)$$

(see Reuter and Ledermann (1953)). We shall then be able to show that the convergence of $p_{ij}(s,s+h)$ to δ_{ij} as $h \to 0$ is uniform with respect to s. Secondly, as was shown by Isaacson and Arnold (1978), in the case of continuous time homogeneous Markov chains, if the intensities of passage become close to zero or grow without bound, it would be impossible to control the behavior of the mean visit times.

For nonhomogeneous Markov chains mean visit times are defined in the following manner. Suppose at time s the chain occupies state i. Let N(i,j,s) be the random variable which counts the number of steps the chain makes until the first visit to state j from state i. We let $T_s(k)$ represent the random variable which measures the time spent between the (k-1)st and kth move for the chain, starting at time s. Thus

$$\mu_{ij}(s) = E\left(\sum_{k=1}^{N(i,j,s)} T_s(k) \right)$$

represents the mean visit time from state i to state j starting at time s.

The main result of this chapter shows that X(t) is uniformly strongly ergodic if and only if the chain is uniformly ergodic and the mean visit times to some state are bounded over the starting states and starting times. The method of proof is similar to the one given by Isaacson and Arnold. We first discretize X(t). Then we characterize strong ergodicity for X(t) in terms of the discretized chains. We use the results of Huang and Isaacson (1976) to relate uniform strong ergodicity for the discretized chains in terms of the behavior of the mean visit times for the discretized chains. Finally, by comparing the mean visit times of the discrete time chains to those of the continuous time chain we are able to prove the desired result.

We discretize X(t) in the following manner. Let Δt be a rational number such that $0<\Delta t\leq 1$. Let $0\leq s\leq 1$. Starting at time s observe X(t) only at times $s+k\cdot\Delta t$ for $k = 0,1,2\ldots$. We then can view this procedure as looking at a discrete time nonhomogeneous Markov chain $Z_{s,\Delta t}(n)$. The transition matrices of $Z_{s,\Delta t}(n)$ are given by the collection

$$\{P(s+k\cdot\Delta t,\ s+(k+1)\cdot\Delta t)\}_{k=0}^{\infty}$$

The first result shows that the collection given immediately above is equicontinuous in s.

To this end consider:

Lemma 6.1:

$$\lim_{h\to 0}||P(s,s+h)-I|| = 0 \quad \text{uniformly in } s.$$

Proof:

$$||P(s,s+h)-I|| = \sup_i \sum_{j=0}^{\infty} |p_{ij}(s,s+h)-\delta_{ij}|$$

$$= \sup_i (\sum_{j\neq i} p_{ij}(s,s+h)+|p_{ii}(s,s+h)-1|)$$

$$= \sup_i (1-p_{ii}(s,s+h) + |p_{ii}(s,s+h)-1|)$$

$$= 2 \sup_i (|1-p_{ii}(s,s+h)|)$$

From (6.3) the integrated form of the solution to the Kolmogorov differential equations was given by

$$p_{ii}(s,s+h) = 1 + \int_s^{s+h} \sum_{k=0}^{\infty} [p_{ik}(s,u)q_{kj}(u)]du$$

Having $|q_{ii}(u)| \leq B$ for all i and u implies that $\sum_{j\neq i} q_{ij}(u) = -q_{ii}(u) = |q_{ii}(u)| \leq B$. Since $q_{ij}(u) \geq 0$ for $i\neq j$ we now can see that $q_{ij}(u) \leq B$ for $i\neq j$. Hence, for all i, j and u $|q_{ij}(u)| \leq B$. Thus

$$|1-p_{ii}(s,s+h)| \leq \int_s^{s+h} \left(\sum_{k=0}^{\infty} p_{ik}(s,u)\cdot B\right)du$$

$$= B\int_s^{s+h} \left(\sum_{k=0}^{\infty} p_{ik}(s,u)\right)du$$

$$= \int_s^{s+h} Bdu = hB$$

Thus $|1-p_{ii}(s,s+h)| \leq h\cdot B$ for all i in S and all $s\geq 0$. Thus

$$||P(s,s+h)-I|| = 2\cdot\sup_i|1-p_{ii}(s,s+h)| \leq 2h\cdot B$$

which goes to zero as h goes to zero uniformly in s. This completes the proof.

From the above lemma we have the following corollary.

Corollary 6.1:

For any s<t, $\lim_{h\to 0}||P(s,s+t)-P(s+h,s+h+t)|| = 0$, uniformly in s and t.

Proof:

Choose h such that h<t. Now

$$||P(s,s+t)-P(s+h,s+h+t)|| = ||P(s,s+t)-P(s+h,s+t)$$

$$+ P(s+h,s+t)-P(s+h,s+h+t)||$$

$$\leq ||P(s,s+t)-P(s+h,s+t)|| + ||P(s+h,s+t)-P(s+h,s+h+t)||$$

$$= ||P(s,s+h)\cdot P(s+h,s+t)-P(s+h,s+t)|| + ||P(s+h,s+t)$$

$$- P(s+h,s+t)P(s+t,s+t+h)||$$

$$= ||(P(s,s+h)-I)\cdot P(s+h,s+t)|| + ||P(s+h,s+t)\cdot$$

$$\cdot\ (I-P(s+t,s+t+h)||$$

$$\leq ||P(s,s+h)-I||\cdot||P(s+h,s+t)|| + ||P(s+h,s+t)||\cdot$$

$$\cdot||P(s+t,\ s+t+h)-I||$$

$$= ||P(s,s+h)-I|| + ||P(s+t,s+t+h)-I||.$$

The last two terms of the above inequality converge to zero, uniformly in s and t, as $h \to 0$ from Lemma 6.1. Hence we have

$$\lim_{h\to 0}||P(s,s+t)-P(s+h,s+t+h)|| = 0$$

uniformly in s and t.

As a consequence of the above corollary we have that the function $p_{ij}(s,s+t)$ is uniformly continuous in s, since

$$|p_{ij}(s+h,s+t+h) - p_{ij}(s,s+t)| \leq ||P(s+h,s+t+h)$$

$$- P(s,s+t)||$$

for all $h>0$. Also the family $\{P(s+kt,s+kt+nt)\}_{n,k=0}^{\infty}$ when viewed as a function of s, is equicontinuous. That is to say, for each $\varepsilon>0$ and all s there exists an h such that for all x with $s-h<x<s+h$,

$$\sup_{n,k,t} ||P(s+kt,s+kt+nt)-P(x+kt,x+kt+nt)|| < \varepsilon.$$

We shall investigate the behavior of $P(s+kt, s+kt+nt)$ as $n\to\infty$, when $0\leq s\leq 1$ and $0<t\leq 1$, t rational. To aid in the investigation we need the following lemma.

Lemma 6.2:

Suppose $\{f_{n,k}(s)\}_{n=1,k=0}^{\infty \quad \infty}$ is an equicontinuous family of functions, where $0\leq s\leq 1$. If for each s

$$\lim_{n\to\infty} f_{n,k}(s) = g_k(s)$$

uniformly in k, then $g_k(s)$ is continuous and the convergence in uniform with respect to s.

Proof:

Let $\varepsilon>0$ be given. By equicontinuity there is an open interval containing s such that

$$||f_{n,k}(s)-f_{n,k}(y)|| < \varepsilon$$

for all y in the interval and for all n and k. Thus

$$||g_k(y)-g_k(s)|| \leq ||g_k(y)-f_{n,k}(y)||$$

$$+ ||f_{n,k}(y)-f_{n,k}(s)|| + ||f_{n,k}(s)-g_k(s)||.$$

From the above the middle term on the right hand side may be made less than ε for all n and k. The outside terms go to zero uniformly in k as $n\to\infty$. Thus

$$||g_k(s)-g_k(y)| < \varepsilon$$

for all y in the interval containing s. So $g_k(s)$ is continuous.

To show that $f_{n,k}(s)$ converges to $g_k(s)$ uniformly in s we need to show that for all $\varepsilon>0$ there exists an $N=N(\varepsilon)$ such that for all $n\geq N$

$$||f_{n,k}(s)-g_k(s)|| < \varepsilon$$

for all s. We know that for $0\leq s\leq 1$ there is an open interval I_s such that

$$||f_{n,k}(s)-f_{n,k}(x)|| < \frac{\varepsilon}{3}$$

for all x in I_s and for all k and n. By the first part of this lemma

$$||g_k(s)-g_k(x)|| < \frac{\varepsilon}{3}$$

since the interval [0,1] is compact, there exists a finite collection of such intervals, $\{I_{s_1},\ldots I_{s_m}\}$, which cover [0,1]. Choose N sufficiently large so that

$$||f_{n,k}(s_i)-g_k(s_i)|| < \frac{\varepsilon}{3}$$

for each s_i and $n\geq N$, and all k. Now for any $0\leq y\leq 1$ there is an i such that y is an element of I_{s_i}. Hence

$$||f_{n,k}(y)-g_k(y)|| \leq ||f_{n,k}(y)-f_{n,k}(s_i)||$$

$$+ ||f_{n,k}(s_i)-g_k(s_i)|| + ||g_k(s_i)-g_k(y)||$$

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,$$

for $n \geq N$ and all k. Thus $f_{n,k}(s)$ converges to $g_k(s)$ uniformly in s as well as k. This completes the proof.

We now apply the latest corollary and lemma to the discretized chains. Suppose that $0 \leq s \leq 1$ and $0 < t \leq 1$, t rational, and there exists a row constant stochastic matrix L(s,t) such that

$$\lim_{n\to\infty} ||P(s+kt,\ s+kt+nt) - L(s,t)|| = 0$$

uniformly in k and s. Thus, by definition, $Z_{s,t}(n)$ is uniformly strongly ergodic. It follows from Corollary 6.1 that for each t, $\{P(s+kt,\ s+kt+nt)\}_{k,n=1}^{\infty}$ is an equicontinuous family. By Lemma 6.2 L(s,t) is continuous in s for each t, and the convergence of P(s+kt,s+kt+nt) to L(s,t) as $n\to\infty$ is uniform with respect to s as well as k. We shall now show that L(s,t) does not depend on s or t.

<u>Lemma 6.3</u>:

If $Z_{s,t}(n)$ is uniformly strongly ergodic with limiting distribution L(s,t), where $0 \leq s \leq 1$, $0 < t \leq 1$, t rational, then L(s,t) is independent of t.

Proof:

By assumption

$$\lim_{n\to\infty} ||P(s+kt,s+kt+nt)-L(s,t)|| = 0$$

uniformly in k. From the comments after Lemma 6.2, the convergence is uniform with respect to s as well as k. Suppose then $\lim_{n\to\infty} P(s,s+nt) = L(s,t)$ uniformly in s. For any subsequence $\{n_k\}$ of $\{n\}$, where $n_k\to\infty$ as $k\to\infty$, we have

$$\lim_{k\to\infty} P(s,s+n_k\cdot t) = L(s,t).$$

Since t is rational we shall choose the subsequence $\{n_k\}$ in the following manner.

$$n_1 = \min\{n: nt \text{ is an integer}\}$$

$$n_2 = \min\{n: nt \text{ is an integer, } n>n_1\}$$

$$\vdots$$

$$n_k = \min\{n: nt \text{ is an integer, } n>n_{k-1}\}$$

$$\vdots$$

Using this subsequence we have

$$\lim_{k\to\infty} P(s,s+n_k\cdot t) = L(s,t)$$

uniformly in s. By assumption we have, when t=1,

$$\lim_{n\to\infty} P(s,s+n) = L(s,1).$$

Again along any subsequence $\{m_k\}$ of $\{n\}$

$$\lim_{k\to\infty} P(s,s+m_k) = L(s,1).$$

Let $m_k = n_k \cdot t$ and hence

$$L(s,t) = \lim_{k\to\infty} P(s,s+n_k \cdot t) = \lim_{k\to\infty} P(s,s+m_k) = L(s,1).$$

Thus $L(s,t) = L(s,1)$ for all t rational, with $0<t\leq 1$. Hence $L(s,t)$ is independent of t, hence we shall write $L(s,t) = L(s)$. This completes this lemma.

We now show that $L(s)$, given above, is independent of s.

<u>Lemma 6.4</u>:

With the same assumptions given in Lemma 6.3, $L(s,t) = L(s) = L$, that is $L(s)$ is independent of s.

<u>Proof</u>:

From Lemma 6.3 $L(s,t) = L(s)$ for all rational t, $0<t\leq 1$. By assumption $\lim_{n\to\infty} P(0,nt) = L(0)$ for all rational t, $0<t\leq 1$. First let s be a rational number with $s = \frac{p}{q}$, where p and q are integers such that $p\leq q$ and $q\neq 0$. Again by assumption

$$\lim_{n\to\infty} P(p/q,\ p/q + nt) = L(p/q).$$

Since t is rational, $t = \frac{p'}{q'}$ and $s + nt = (pq'+p'q\cdot n) \cdot \frac{1}{q\cdot q'}$. Since $1/qq'$ is rational

$$\lim_{n\to\infty} P(0,n\cdot \frac{1}{qq'}) = L(0).$$

Let $m_n = pq' + n\cdot p'q$ and by the argument given in Lemma 6.3

$$L(0) = \lim_{n\to\infty} P(0,m_n\cdot \frac{1}{qq'}) = \lim_{n\to\infty} P(0,s+nt).$$

Consider

$$||P(0,s+nt) - L(s)|| = ||P(0,s)P(s,s+nt)-L(s)||$$

$$= ||P(0,s)\cdot P(s,s+nt)-P(0,s)\cdot L(s)||$$

$$= ||P(0,s)(P(s,s+nt)-L(s))||$$

$$\leq ||P(0,s)|| \cdot ||P(s,s+nt)-L(s)||$$

$$= ||P(s,s+nt)-L(s)||.$$

Since $\lim_{n\to\infty} P(s,s+nt) = L(s)$ we have that $\lim_{n\to\infty} P(0,s+nt) = L(s)$. Yet we showed that $\lim_{n\to\infty} P(0,s+nt) = L(0)$, hence $L(s) = L(0)$ for s rational and $0\leq s\leq 1$.

Now we show, by using the continuity of $L(s,t)$, that $L(s) = L(0)$ for all $0\leq s\leq 1$. For an irrational number r with $0<r<1$, let s_n be a sequence of rationals such that $\lim_{n\to\infty} s_n = r$. Since $L(s,t) = L(s)$ is continuous for all s, $\lim_{n\to\infty} L(s_n) = L(r)$. Yet $L(s_n) = L(0)$ for all n. Hence $L(r) = L(0)$ for all $0<r<1$, r irrational. Hence for all s,

$0\leq s\leq 1, L(s) = L(0)$. Therefore $L(s)$ is independent of s, hence we shall write $L(s) = L$.

Combining these results we have the following theorem.

Theorem 6.1:

If $Z_{s,t}(n)$ is uniformly strongly ergodic for all $0\leq s\leq 1$ and $0<t\leq 1$, t rational, then the limiting distribution, $L(s,t)$, does not depend on s or t. Hence we shall write $L(s,t) = L$.

We are now ready to characterize strong ergodicity for $X(t)$ in terms of strong ergodicity for the discretized chains.

Theorem 6.2:

$X(t)$ is uniformly strongly ergodic if and only if for all $0\leq s\leq 1$ and $0<t\leq 1$, t rational, $Z_{s,t}(n)$ is uniformly strongly ergodic.

Proof:

Suppose $X(t)$ is uniformly strongly ergodic. Then there exists a row constant stochastic matrix L such that $\lim_{t\to\infty}||P(s,s+t)-L|| = 0$, uniformly in s. Thus given $\varepsilon>0$ there exists a $T = T(\varepsilon)$, such that for $t>T$

$$||P(s,s+t)-L|| < \varepsilon,$$

for all s. Let t be rational and $0<t\leq 1$. Let n be an

integer such that $nt>T$. For any $0\leq s\leq 1$ and any integer k, we have

$$||P(s+kt,s+kt+nt)-L|| < \varepsilon.$$

Hence $Z_{s,t}(n)$ is uniformly strongly ergodic.

Conversely, suppose for $0\leq s\leq 1$ and $0<t\leq 1$, t rational, that $Z_{s,t}(n)$ is uniformly ergodic. By Theorem 6.1 the limiting distribution of $Z_{s,t}(n)$ does not depend on s or t. Hence there exists a row constant stochastic matrix L, such that

$$\lim_{n\to\infty}||P(s+kt,s+kt+nt)-L|| = 0$$

uniformly in k. By Lemma 6.2 the convergence is uniform with respect to s. Hence for all $\varepsilon>0$ there exists an $N = N(\varepsilon,t)$ such that for $n>N$

$$||P(s+kt,s+kt+nt)-L|| < \varepsilon,$$

for all s and k. Let u and v be reals such that $v>N+2$. Let $m_1 = \min\{k: kt\geq u, k \text{ an integer}\}$ and $m_2 = \max\{k: m_1t + kt\leq u+v\}$. Since $0<t\leq 1$, $u+v-1<(m_1+m_2)t$. If not, then $u+v\geq(m_1+m_2)t + 1\geq(m_1+m_2)t+t = (m_1+m_2+1)t$. This contradicts the assumption that m_2 is the maximum integer such that $(m_1+m_2)t \leq u+v$. Hence $u+v-1 < (m_1+m_2)t \leq u+v$. Hence there exists an r such that $(m_1+m_2)t + r = u+v$. Suppose $m_2\cdot t > v$, then $(m_1+m_2)t > v+m_1t \geq v+u$, again a contradiction. Hence $m_2\cdot t \leq v$. Also $m_2t > v-2$. Otherwise, we would have $v>2+m_2\cdot t = (m_2t+1) + 1 \geq (m_2t+1) + m_1t-u \geq m_2t + t + m_1t - u$. Thus $v + u > (m_1+m_2)t + t$, a contradiction. Therefore

$$m_2 \geq m_2t > v-2 \geq N+2 - 2 = N \text{ or } m_2 > N.$$

Combining these results we have that

$$||P(u,u+v)-L|| = ||P(u,(m_1+m_2)t+r)-L||$$

$$= ||P(u,r+m_1t)P(r+m_1t,r+(m_1+m_2)t-L||$$

$$= ||P(u,r+m_1t)P(r+m_1t,r+(m_1+m_2)t)-P(u,r+m_1t)L||$$

$$= ||P(u,r+m_1t)(P(r+m_1t,r+(m_1+m_2)-L)||$$

$$\leq ||P(u,r+m_1t)||\cdot||P(r+m_1t,r+(m_1+m_2)t)-L||$$

$$= ||P(r+m_1t,r+(m_1+m_2)t)-L||.$$

By assumption, since $m_2>N$, $||P(r+m_1t,\ r+m_1t+m_2t)-L|| < \varepsilon$. Hence there exists a $T=N+2$, such that for $v>T$

$$||P(u,u+v)-L|| < \varepsilon$$

uniformly in u. Hence X(t) is uniformly strongly ergodic. This completes the proof.

We now shall incorporate the notions of mean visit times. Let Δt be a fixed rational such that $0<\Delta t\leq 1$. Let $0\leq s\leq 1$. Define the probability of a first visit to state j at time k+n given that at time k the chain $Z_{s,\Delta t}(n)$ occupied state i as ${}_sf_{ij}(k,k+n) = P(Z_{s,\Delta t}(k+n) = j,\ Z_{s,\Delta t}(k+n-1) \neq j, \ldots, Z_{s,\Delta t}(k+1) \neq j \mid Z_{s,\Delta t}(k)=i)$. Define ${}_sm_{ij}(k) = \sum_{m=1}^{\infty} m \cdot {}_sf_{ij}(k,k+m)$. Thus ${}_sm_{ij}(k)$ is the mean visit time from state i to state j, starting at time k, using the transitions of $Z_{s,\Delta t}(n)$.

For a discrete time nonhomogeneous Markov chain, uniform strong ergodicity implies having the mean visit times to some positive recurrent state bounded over the starting states and starting times (Huang and Isaacson, 1976). Hence if $Z_{s,\Delta t}(n)$ is uniformly strongly ergodic, and there exists at least one positive recurrent state, 0, say, then $\sup_{i,k} {}_sm_{i0}(k) < +\infty$. We now show that the supremum may be taken over all s.

Theorem 6.3:

Suppose $Z_{s,\Delta t}(n)$ is uniformly strongly ergodic for each s and each rational Δt with $0<\Delta t\leq 1$. If we fix Δt, then $\sup_{s,i,k} {}_sm_{i0}(k) < \infty$.

Proof:

From Theorem 6.1

$$\lim_{n\to\infty} P(s+k\Delta t, s+k\Delta t+n\Delta t) = L$$

uniformly in k and s. Hence

$$\lim_{n\to\infty} p_{ii}(s+k\Delta t, s+k\Delta t+n\Delta t) = \pi_i \geq 0,$$

uniformly in k and s, with $\sum_{i=0}^{\infty} \pi_i = 1$. Since state 0 is positive recurrent $\pi_0 > 0$. Now there exists an integer $N = N(\pi_0)$ such that

$$||P(s+k\Delta t, s+k\Delta t+n\Delta t) - L|| < \frac{\pi_0}{2}$$

for all $n \geq N$, for all s, and all k. Hence

$$|p_{i0}(s+k\Delta t, s+k\Delta t+n\Delta t) - \pi_0| < \frac{\pi_0}{2}$$

for all s, k, i and $n \geq N$.

For any $r = 0,1,2,\ldots,N-1$ we have

$$\begin{aligned} {}_sf_{i0}(k,k+2N+r) &\leq \sum_{\ell\neq 0} p_{i\ell}(s+k\Delta t, s+(k+N)\Delta t) \\ &\quad \cdot {}_sf_{\ell 0}(k+N,k+2N+r) \\ &\leq \sum_{\ell\neq 0} p_{i\ell}(s+k\Delta t, s+(k+N)\Delta t) \\ &= 1 - p_{i0}(s+k\Delta t, s+(k+N)\Delta t) \\ &< 1 - \frac{\pi_0}{2} \end{aligned}$$

for all i, s, k. Similarly,

$${}_sf_{i0}(k,k+3N+r) \leq \sum_{\ell\neq 0} p_{i\ell}(s+k\Delta t, s+(k+N)\Delta t)\, {}_sf_{\ell 0}(k+N, k+3N+r).$$

From the above argument ${}_sf_{i0}(m,m+2N+r) < (1 - \frac{\pi_0}{2})$ for all i, s, m and $r = 0,1,2,\ldots,N-1$. For $m = k+N$ we have

${}_s f_{\ell 0}(k+N, k+3N+r) < 1 - \frac{\pi_0}{2}$ for all ℓ, s, and $r = 0,1,2,\ldots,$ $N-1$. Hence

$$ {}_s f_{i0}(k, k+3N+r) < \left(1 - \frac{\pi_0}{2}\right) \sum_{\ell \neq 0} p_{i\ell}(s+k\Delta t, s+(k+N)\Delta t) $$

$$ < \left(1 - \frac{\pi_0}{2}\right)^2. $$

Continuing in this manner

$$ {}_s f_{i0}(k, k+n\cdot N+r) < \left(1 - \frac{\pi_0}{2}\right)^{n-1} $$

for all i, s, k, $n=2,3,4,\ldots,$ and $r=0,1,2,\ldots,N-1$.

Computing ${}_s m_{i0}(k)$ we see that

$$ {}_s m_{i0}(k) = \sum_{n=1}^{\infty} n \cdot {}_s f_{i0}(k, k+n) $$

$$ = \sum_{n=1}^{2N-1} n \cdot {}_s f_{i0}(k, k+n) + \sum_{n=2N}^{3N-1} n \cdot {}_s f_{i0}(k, k+n) $$

$$ + \sum_{n=3N}^{4N-1} n \cdot {}_s f_{i0}(k, k+n) + \ldots $$

$$ \leq \sum_{n=1}^{2N-1} n + \left(1 - \frac{\pi_0}{2}\right) \sum_{n=2N}^{3N-1} n + \left(1 - \frac{\pi_0}{2}\right)^2 \sum_{n=3N}^{4N-1} n + \ldots $$

$$ \leq 2N^2 + \left(1 - \frac{\pi_0}{2}\right) 3N^2 + \left(1 - \frac{\pi_0}{2}\right)^2 \cdot 4N^2 + \ldots $$

$$ = N^2 \cdot \sum_{n=2}^{\infty} n \cdot \left(1 - \frac{\pi_0}{2}\right)^{n-2} < \infty. $$

Note that the bound on ${}_s m_{i0}(k)$ is independent of s, i, and k. Therefore $\sup_{i,s,k} {}_s m_{i0}(k) < \infty$. Hence we are done.

We are now ready to prove the main theorem of this section, by comparing the mean visit times for X(t) to those of $Z_{s,t}(n)$.

Theorem 6.4:

X(t) is uniformly strongly ergodic if and only if X(t) is uniformly ergodic and $\sup_{i,u} \mu_{i0}(u) < \infty$.

Proof:

Assume X(t) is uniformly strongly ergodic. By Theorem 6.2 for all $0 \leq s \leq 1$ and $0 < t \leq 1$, t rational, $Z_{s,t}(n)$ is uniformly strongly ergodic. Fix Δt rational with $0 < \Delta t \leq 1$. By Theorem 6.3 $\sup_{i,s,k} {}_{s}m_{i0}(k) < +\infty$. Let u be any starting time for X(t). We may write $u = k\Delta t + r$, where $0 \leq r \leq 1$ and k is an integer. Since X(t) may visit state zero at a time other than times of the form $r+k\Delta t+n\Delta t$ we have

$$\mu_{i0}(u) = \mu_{i0}(r+k\Delta t) \leq {}_{r}m_{i0}(k) \leq \sup_{i,r,k} {}_{r}m_{i0}(k) < \infty.$$

Thus

$$\sup_{i,u} \mu_{i0}(u) < \infty.$$

Conversely assume that $\sup_{i,u} \mu_{i0}(u) < \infty$. Pick s and Δt such that $0 \leq s \leq 1$, $0 < \Delta t \leq 1$, and Δt is rational. Let $Z_{s,\Delta t}(n)$ be the discrete time nonhomogeneous chain as given earlier. From Theorem 6.2 we know that X(t) is uniformly strongly ergodic if and only if $Z_{s,\Delta t}(n)$ is uniformly strongly ergodic for all s and Δt. We shall show that for each Δt, $\sup_{i,s,k} {}_s m_{i0}(k) < +\infty$. Since we assumed that $X(t)$ is uniformly ergodic, $Z_{s,\Delta t}(n)$ will be uniformly strongly ergodic for each s and Δt. (See Huang and Isaacson (1976)).

Since $X(t)$ may visit zero several times before $Z_{s,\Delta t}(n)$ visits zero further notation is needed. Let u be any starting time for $X(t)$, and again $u = r+h\cdot\Delta t$, where h is an integer and $r \leq \Delta t \leq 1$. Define

${}_u W_k(0)$ = the waiting time from when $X(t)$ leaves zero for the kth time, until $X(t)$ returns to zero, starting at time u.

${}_u W(i)$ = the waiting time until the first visit to zero from state i by $X(t)$, starting at time u.

${}_u D_k(0)$ = the time that $X(t)$ spends in state zero during the kth visit, starting at time u.

${}_u W^*(i)$ = the waiting time to visit state zero from state i at time $u = r+h\Delta t$, using the transitions of $Z_{r,\Delta t}(n)$

Even though the ${}_u D_k(0)$'s are not identically distributed, we still have that for each Δt, $P({}_u D_k(0) \leq \Delta t) \leq \beta = \beta(\Delta t) < 1$ for all k and u. This is possible since the behavior of the distribution of the time to leave state zero is governed by the behavior of

$$\exp\left[\int_t^{t+\Delta t} q_{00}(s)\,ds\right],$$

which is contained in the interval $(e^{-B\Delta t} \leq x \leq e^{-b\Delta t}) \subseteq (0<x<1)$, for all t.

Let M_u be a random variable such that $M_u = m$ if the mth visit to zero by X(t) is the first visit to zero by $Z_{s,\Delta t}(n)$, starting at time $u = r+h\Delta t$. Hence, we may write

$$ {}_uW^*(i) \leq {}_uW(i) + \sum_{j=0}^{M_u-1} ({}_uW_j(0)+{}_uD_j(0)) + \Delta t, $$

where ${}_uW_0(0) = {}_uD_0(0) = 0$. Thus

$$ {}_rm_{i0}(h) = E({}_uW^*(i)) = \sum_{k=0}^{\infty} P({}_uW^*(i)>k) $$

$$ \leq \sum_{k=0}^{\infty} P({}_uW(i) + \sum_{j=0}^{M_u-1} ({}_uW_j(0)+{}_uD_j(0))>k) + \Delta t $$

$$ = \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} P({}_uW_{(i)} + \sum_{j=0}^{m-1} ({}_uW_j(0)+{}_uD_j(0))>k, M_u=m)+\Delta t $$

If $M_u=m$ where $u = r+h\Delta t$, then the time spent in state zero before leaving during the jth visit must be less than Δt for $j = 0,1,2,\ldots m-1$. Thus

$$ [M_u=m] \subseteq (\bigcap_{j=1}^{m-1} ({}_uD_j(0)\leq\Delta t)) $$

Hence

$$ E({}_uW^*(i)) \leq \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} P({}_uW(i) + \sum_{j=0}^{m-1} ({}_uW_j(0)+{}_uD_j(0)) > k, $$

$$ \bigcap_{j=1}^{m-1} ({}_uD_j(0)<\Delta t)) + \Delta t $$

$$ \leq \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} P({}_uW(i) + \sum_{j=0}^{m-1} {}_uW_j(0) + m\Delta t > k, $$

$$ \bigcap_{j=1}^{m-1} ({}_uD_j(0)<\Delta t) + \Delta t $$

From the Markov property the ${}_uW_j(0)$'s and ${}_uW(i)$ are stochastically independent of the ${}_uD_j(0)$'s. Also the ${}_uD_j(0)$'s are independent random variables for $j = 1,2,\ldots$. Thus

$$E({}_uW^*(i)) \leq \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} P({}_uW(i) + \sum_{j=0}^{m-1} {}_uW_j(0)$$

$$+ m\Delta t>k) \cdot P(\bigcap_{j=1}^{m-1} ({}_uD_j(0)<\Delta t)) + \Delta t$$

$$= \sum_{k=0}^{\infty} \sum_{m=1}^{\infty} P({}_uW(i) + \sum_{j=0}^{m-1} {}_uW_j(0) + m\Delta t>k)$$

$$\cdot \prod_{j=1}^{m-1} P({}_uD_j(0)<\Delta t) + \Delta t$$

$$\leq \sum_{m=1}^{\infty} \sum_{k=0}^{\infty} P({}_uW(i) + \sum_{j=0}^{m-1} {}_uW_j(0)+m\Delta t>k) \cdot \beta^{m-1} + \Delta t$$

$$= \sum_{m=1}^{\infty} (E({}_uW(i) + \sum_{j=0}^{m-1} {}_uW_j(0))+m\Delta t) \cdot \beta^{m-1} + \Delta t$$

$$\leq \sum_{m=1}^{\infty} (\sup_i \mu_{i0}(u)+(m-1)\sup_u \mu_{00}(u) + m\Delta t) \cdot \beta^{n-1} + \Delta t$$

$$\leq \sum_{m=1}^{\infty} m\cdot(\sup_{i,u} \mu_{i0}(u)+\Delta t) \cdot \beta^{m-1} + \Delta t <\infty$$

for each Δt. Hence

$${}_r m_{i0}(h) \leq \sum_{m=1}^{\infty} m\cdot(\sup_{i,u}(\mu_{i0}(u))+\Delta t) \cdot \beta^{m-1} + \Delta t$$

for all r, h and i, since $\beta<1$. Thus

$$\sup_{r,i,h} {}_r m_{i0}(h) < +\infty$$

for each Δt. Hence for each r and Δt $Z_{r,\Delta t}(n)$ is uniformly strongly ergodic. Thus $X(t)$ is uniformly strongly ergodic, by Theorem 6.2. This completes the proof.

VII. STRONG ERGODICITY FOR CONTINUOUS TIME NONHOMOGENEOUS CHAINS USING A RELATED DISCRETE TIME CHAIN

From the work done in Chapter VI, the characterization of strong ergodicity for continuous time nonhomogeneous chains is much more tedious, analytically, than characterizations for continuous time homogeneous chains. Yet, by imposing regularity conditions on the intensity of passage matrix, we can reduce the mathematics considerably, to obtain strong ergodicity for continuous time nonhomogeneous Markov chains.

For this chapter let $X(t)$ be a continuous time nonhomogeneous Markov chain defined on the state space $S = \{0,1,2,...\}$. Let $P(s,t)$ be the matrix of transition probability functions. Let $Q(s)$ be the transition intensity matrix, where

$$Q(s) = \lim_{h\to 0} \frac{P(s,s+h)-I}{h} .$$

Suppose that $Q(s)$ has the special form

$$Q(s) = Q\cdot h(s).$$

where Q is a matrix such that $\sum_{i\neq j} q_{ij} = -q_{ii}$ for all i. Also $-q_{ii}\geq 0$ for all i and $q_{ij}\geq 0$ for all $i\neq j$. That is, Q is an intensity matrix. The function $h(s)$ is such that

a) $h(s)$ is nonnegative, and $h(t) \to 1$ as $t \to \infty$.

b) $H(t) = \int_0^t h(s)ds$ is strictly increasing.

c) $H(t) \to \infty$ as $t \to \infty$.

Johansen and Ramsey (1978) discussed the use of having $Q(s) = Q \cdot h(s)$ in the study of embedding problems. One other reason to use the special form of the intensity matrix can be explained by the choice of $h(s)$. That is, by having $\lim_{s\to\infty} h(s) = 1$ then $\lim_{s\to\infty} Q(s) = Q$. Hence for large values of s, the intensity of movement is governed by an "almost" homogeneous Markov chain.

We know that the transition probabilities satisfy the following

$$\frac{\partial p_{ij}(s,t)}{\partial t} = \sum_{k=0}^{\infty} p_{ik}(s,t) q_{kj}(t) \tag{7.1}$$

and

$$-\frac{\partial}{\partial s} p_{ij}(s,t) = \sum_{k=0}^{\infty} q_{ik}(s) \cdot p_{kj}(s,t) \tag{7.2}$$

for all i, j, s and t. We now demonstrate that there exists a closed form solution to these equations, when we assume $Q(s) = Q \cdot h(s)$.

<u>Theorem 7.1</u>:

The Kolmogorov differential equations, given above, have as a solution

$$F(s,t) = I + \sum_{n=1}^{\infty} \left(\frac{H(t)-H(s)}{n!}\right)^n Q^n,$$

when the intensity matrix $Q(s) = Q\cdot h(s)$ and $\sup_i |q_{ii}| = q<\infty$.

Proof:

All that needs to be shown is that the partials of each element of $F(s,t)$ satisfy (7.1) and (7.2). So, consider for $i\neq j$

$$\frac{\partial}{\partial t} f_{ij}(s,t) = \frac{\partial}{\partial t} \sum_{n=1}^{\infty} \left(\frac{H(t)-H(s)^n}{n!}\right) q_{ij}(n).$$

Where $q_{ij}(n)$ is the (ij)th element of the matrix Q^n. To differentiate the series termwise, it is sufficient to show that the series of partial derivatives converges absolutely. To this end

$$\sum_{n=1}^{\infty} \left|\frac{\partial}{\partial t} \left(\frac{H(t)-H(s)^n}{n!}\right) q_{ij}(n)\right|$$

$$= \sum_{n=1}^{\infty} \left|\frac{n(H(t)-H(s))^{n-1}\cdot h(t)}{n!} q_{ij}(n)\right|$$

$$= \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!}|q_{ij}(n)|\cdot h(t).$$

Since $\sup_i |q_{ii}| = q < +\infty$,

$$||Q|| = \sup_i \sum_j |q_{ij}| = \sup_i (\sum_{j\neq i} |q_{ij}| + -q_{ii})$$

$$= \sup_i (\sum_{j\neq i} q_{ij} + -q_{ii})$$

$$= \sup_i (-q_{ii} + -q_{ii})$$

$$= 2 \sup_i |q_{ii}| = 2q.$$

Also

$$||Q^n|| \leq ||Q||^n \quad \text{for all } n. \quad \text{Thus}$$

$$|q_{ij}(n)| \leq \sum_{j=0}^{\infty} |q_{ij}(n)| \leq ||Q^n|| \leq ||Q||^n = (2q)^n.$$

Replacing $|q_{ij}(n)|$ by its bound in the series of partial derivatives, we have

$$\sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} |q_{ij}(n)| \cdot h(t)$$

$$\leq h(t) \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} (2q)^n$$

$$= h(t) \cdot 2q \cdot e^{(H(t)-H(s))2q}$$

Thus the series of partial derivatives is absolutely convergent. Hence

$$\frac{\partial}{\partial t} f_{ij}(s,t) = \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} \cdot q_{ij}(n) \cdot h(t)$$

Now

$$q_{ij}(n) = \sum_{k=0}^{\infty} q_{ik}(n-1)\cdot q_{kj}, \text{ so}$$

$$\frac{\partial}{\partial t} f_{ij}(s,t) = \sum_{n=1}^{\infty} \left[\frac{(H(t)-H(s))^{n-1}}{(n-1)!} h(t) \left(\sum_{k=0}^{\infty} q_{ik}(n-1) q_{kj}\right)\right]$$

$$= \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} h(t)\cdot q_{ik}(n-1)\cdot q_{kj}.$$

Consider

$$\sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} h(t)\cdot |q_{ik}(n-1)|\cdot|q_{kj}|$$

$$\leq \sum_{n=1}^{\infty} \sum_{k=0}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} h(t)\cdot |q_{ik}(n-1)|\cdot q$$

$$= \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} h(t)\cdot q(2q)^{n-1} < +\infty$$

Hence we may interchange the order of summation by Fubini's Theorem. So

$$\frac{\partial}{\partial t} f_{ij}(s,t) = \sum_{k=0}^{\infty} q_{kj} h(t)\cdot \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^{n-1}}{(n-1)!} q_{ik}(n-1)$$

$$= \sum_{k=0}^{\infty} q_{kj} h(t)\cdot \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^{n}}{n!} q_{ik}(n)$$

$$= \sum_{k=0}^{\infty} q_{kj} h(t)\cdot f_{ik}(s,t)$$

$$= \left(\sum_{k=0}^{\infty} f_{ik}(s,t) q_{kj}\right) h(t)$$

Thus $f_{ij}(s,t)$ satisfies the forward equation. In a similar manner we show that $-\frac{\partial}{\partial s} f_{ij}(s,t) = \sum_{k=0}^{\infty} h(s) \cdot q_{ik} f_{kj}(s,t)$. Hence $F(s,t)$ solves the Kolmogorov differential equations. This completes the proof.

Even though $F(s,t)$ is a solution to the equations, there is no immediate guarantee that $F(s,t) = P(s,t)$. Reuter and Ledermann (1953) demonstrated that a sufficient condition for the solution to be unique is to have for every $s \leq t$ and for some i

$$\sum_{j=0}^{\infty} f_{ij}(s,t) = 1,$$

that is for some row of $F(s,t)$ $(f_{ij}(s,t))_{j=0}^{\infty}$ forms a probability distribution on S. In our case for any $s \leq t$ and any i

$$\sum_{j=0}^{\infty} f_{ij}(s,t) = \sum_{j=0}^{\infty} (\delta_{ij} + \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^n}{n!} q_{ij}(n)$$

$$= 1 + \sum_{j=0}^{\infty} \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^n}{n!} q_{ij}(n)$$

By an argument similar to the one given in Theorem 7.1 we may interchange the order of summation. Thus

$$\sum_{j=0}^{\infty} f_{ij}(s,t) = 1 + \sum_{n=1}^{\infty} \sum_{j=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} q_{ij}(n)$$

$$= 1 + \sum_{n=1}^{\infty} [\frac{(H(t)-H(s))^n}{n!} \sum_{j=0}^{\infty} q_{ij}(n)]$$

The intensity matrix Q has rows sums equal to zero. We can show that the sum of the elements of any row of Q^n, $m \geq 1$, equals zero by induction. If $\sum_{j=0}^{\infty} q_{ij}(n) = 0$ for all i then consider the sum of the elements of the ith row of Q^{n+1}, that is consider

$$\sum_{j=0}^{\infty} q_{ij}(n+1) = \sum_{j=0}^{\infty} \sum_{k=0}^{\infty} q_{ik} q_{kj}(n)$$

$$= \sum_{k=0}^{\infty} q_{ik} \cdot \sum_{j=0}^{\infty} q_{kj}(n)$$

$$= \sum_{k=0}^{\infty} q_{ik} \cdot 0 = 0.$$

Hence the row sums of Q^{n+1} equal zero. Thus the sum of any row of F(s,t) is 1, since

$$\sum_{j=0}^{\infty} f_{ij}(s,t) = 1 + \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^n}{n!} \sum_{j=0}^{\infty} q_{ij}(n)$$

$$= 1 + \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} \cdot 0 = 1.$$

Therefore the solution to the Kolmogorov differential equations is unique. Thus F(s,t) = P(s,t), or

$$P(s,t) = I + \sum_{n=1}^{\infty} \frac{(H(t)-H(s))^n}{n!} Q^n.$$

Hence given $Q(s) = Q.h(s)$ we can determine P(s,t).

Example 7.1:

Let $Q(s) = \begin{pmatrix} -1 & 1 \\ 2 & -2 \end{pmatrix}\left(\frac{s+2}{s+1}\right) = Q\cdot h(s)$. Clearly

$$H(t) = \int_0^t h(s)ds = t + \ln(t+1) \text{ and}$$

$$Q^n = \begin{bmatrix} 3^{n-1}(-1)^n & 3^{n-1}(-1)^{n+1} \\ 2\cdot 3^{n-1}(-1)^{n+1} & 2\cdot 3^{n-1}(-1)^n \end{bmatrix}.$$

Hence

$$P(s,t) = I + \sum_{n=1}^{\infty} \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix} \frac{3^{n-1}(-1)^n (t-s+\ln \frac{t+1}{s+1})^n}{n!}$$

$$P(s,t) = I + \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix} \frac{1}{3} \sum_{n=1}^{\infty} \frac{(-3(t-s+\ln(\frac{t+1}{s+1})))^n}{n!}$$

$$P(s,t) = I + \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix} \frac{1}{3} \cdot \left[e^{-3(t-s)}\left(\frac{s+1}{t+1}\right)^3 - 1\right]$$

$$P(s,t) = \begin{bmatrix} \frac{2}{3} + \frac{1}{3} e^{-3(t-s)}\left(\frac{s+1}{t+1}\right)^3 & \frac{1}{3} - \frac{1}{3} e^{-3(t-s)}\left(\frac{s+1}{t+1}\right)^3 \\ \frac{2}{3} - \frac{2}{3} e^{-3(t-s)}\left(\frac{s+1}{t+1}\right)^3 & \frac{1}{3} + \frac{2}{3} e^{-3(t-s)}\left(\frac{s+1}{t+1}\right)^3 \end{bmatrix}$$

Following the methods used in Chapter III, let c be any real number such that $c>q = \sup_i |q_{ii}|$. And again, define

$$\overline{P} = I + \frac{Q}{c}.$$

By taking powers of $\overline{P}$ we generate a discrete time Markov chain, $\overline{X}(k)$. Thus, $Q = c(\overline{P}-I)$ and hence

$$P(s,t) = I + \sum_{k=1}^{\infty} \frac{(H(t)-H(s))^n c^n (\overline{P}-I)^n}{n!} .$$

From an argument similar to the one given in Chapter III.

$$P(s,t) = e^{-c(H(t)-H(s))} \cdot (I + \sum_{n=1}^{\infty} \frac{c^n (H(t)-H(s))^n}{n!} \overline{P}^n)$$

By exploiting the functional relationship between $P(s,t)$ and $\overline{P}$ we find desirable properties of $X(t)$ through $\overline{X}(n)$. It should be noted that Yong (1976) proved the same results in the homogeneous setting.

Theorem 7.2:

$X(t)$ is irreducible if and only if $\overline{X}(n)$ is irreducible.

Proof:

If $\overline{X}(n)$ is irreducible then for any states i and j there exists an integer n such that $\overline{p}_{ij}(n) > 0$. Now for any $s \leq t$

$$p_{ij}(s,t) = (\delta_{ij} + \sum_{m=1}^{\infty} \frac{c^m (H(t)-H(s))^m}{m!} \overline{p}_{ij}(m))$$

$$\cdot \; e^{-c(H(t)-H(s))}$$

$$\geq \frac{(H(t)-H(s))^n \cdot c^n \cdot \overline{p}_{ij}(n)}{n!} \; e^{-c(H(t)-H(s))} > 0.$$

Thus $X(t)$ is irreducible.

If $X(t)$ is irreducible then for any states i and j with $i \neq j$ there exists an s and t with $s<t$ such that

$p_{ij}(s,t) > 0$. Hence $0 < \sum_{m=1}^{\infty} \frac{c^m \cdot (H(t)-H(s))^m}{m!} \cdot \bar{p}_{ij}(m)$. So there exists an n such that $\bar{p}_{ij}(m) > 0$. Thus $\bar{X}(n)$ is irreducible, and we are done.

Before we come to the question of ergodic behavior, we need the following lemma.

Lemma 7.1:

For each s and each n,

$$\frac{e^{-c(H(t)-H(s))}(H(t)-H(s))^n}{n!} = o(1) \text{ as } t \to \infty.$$

Proof:

Use L'Hospital's rule and the fact that $H(t) \to \infty$.

Suppose that $\bar{X}(k)$ is ergodic. That is $\lim_{n\to\infty} \bar{p}_{ij}(n) = \pi_j$ independent of i for all j, with $\pi_j \geq 0$ and $\sum_{j=0}^{\infty} \pi_j = 1$. Following Yong (1976),

$$e^{-c(H(t)-H(s))} \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n (\bar{p}_{ij}(n) - \pi_j)$$

$$= e^{-c(H(t)-H(s))} [\sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n \bar{p}_{ij}(n)$$

$$- \pi_j \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n]$$

$$= e^{-c(H(t)-H(s))} \cdot \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n \bar{p}_{ij}(n) - \pi_j$$

$$= p_{ij}(s,t) - \pi_j.$$

Since $\overline{X}(n)$ is ergodic, for each i and for all $\varepsilon>0$ there exists an $N = N(\varepsilon,i)$ such that for any $n>N$

$$|\overline{p}_{ij}(n)-\pi_j| < \frac{\varepsilon}{2}.$$

Thus

$$|p_{ij}(s,t)-\pi_j| \le e^{-(H(t)-H(s))c}$$

$$\cdot \sum_{n=0}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n \; |\overline{p}_{ij}(n)-\pi_j|$$

$$\le e^{-(H(t)-H(s))c}\Big[\sum_{n=0}^{N} \frac{(H(t)-H(s))^n}{n!} c^n |\overline{p}_{ij}(n)-\pi_j|$$

$$+ \frac{\varepsilon}{2} \sum_{n=N+1}^{\infty} \frac{(H(t)-H(s))^n}{n!} c^n\Big]$$

$$< e^{-(H(t)-H(s))c} \cdot \sum_{n=0}^{N} \frac{(H(t)-H(s))^n}{n!} c^n \cdot 2 + \frac{\varepsilon}{2}.$$

By Lemma 7.1 for $n = 0,1,2,\ldots N$ we may choose t so large so that $e^{-(H(t)-H(s))c} \dfrac{(H(t)-H(s))^n}{n!} c^n < \dfrac{\varepsilon}{(N+1)4}$

So for this choice of t we have

$$|p_{ij}(s,t)-\pi_j| < \sum_{n=0}^{N} \left(\frac{\varepsilon}{N+1}\right)\frac{1}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

Thus

$$\lim_{t\to\infty} p_{ij}(s,t) = \pi_j.$$

Hence we have the following result.

Theorem 7.3:

If $\overline{X}(k)$ is ergodic then $X(t)$ is ergodic.

Note that we do not have uniform ergodicity. For $\frac{e^{-(H(t)-H(s))}(H(t)-H(s))^n}{n!} c^n$ does converge to zero, but not independently of s.

Theorem 7.4:

If $\overline{X}(k)$ is strongly ergodic then $X(t)$ is strongly ergodic.

Proof:

Mimic the proof of Theorem 7.3 by replacing $\overline{p}_{ij}(n)$ by $\overline{P}^{(n)}$, $p_{ij}(s,t)$ by $P(s,t)$, π_j by the row constant stochastic matrix L, and absolute values by the matrix norm.

Again we will not have $X(t)$ uniformly strongly ergodic.

When a discrete or continuous time homogeneous Markov chain is ergodic, the long run distribution, L, satisfies

$$LP(t) = L$$

for all t. When we have a nonhomogeneous chain that is strongly ergodic we know that

$$\lim_{t\to\infty} ||P(s,t)-L|| = 0$$

for each s. And hence

$$\lim_{t\to\infty} ||L\cdot P(s,t)-L|| = 0.$$

Yet LP(s,t) need not equal L. For the nonhomogeneous chains we are investigating in this section we do have that

$$LP(s,t) = L.$$

Lemma 7.2:

If $\bar{X}(k)$ is ergodic, with stationary matrix L, then $LP(s,t) = L$ for all $s \leq t$.

Proof:

$$LP(s,t) = L\cdot\left(I + \sum_{j=1}^{\infty} \frac{c^j(H(t)-H(s))^j}{j!}\bar{P}^{(j)}\right)e^{-c(H(t)-H(s))}$$

$$= \left(L + \sum_{j=1}^{\infty} \frac{c^j(H(t)-H(s))^j}{j!} L\bar{P}^{(j)}\right)e^{-c(H(t)-H(s))}$$

$$= \left(L + \sum_{j=1}^{\infty} \frac{c^j(H(t)-H(s))^j}{j!} L\right)e^{-c(H(t)-H(s))}$$

$$= L\cdot e^{c(H(t)-H(s))} \cdot e^{-c(H(t)-H(s))}$$

$$= L.$$

VIII. REFERENCES

Chung, K. 1967. Markov Chains with Stationary Transition Probabilities, 2nd ed. Springer-Verlag, Berlin.

Cinlar, E. 1975. Introduction to Stochastic Processes. Prentice-Hall, Englewood Cliffs, N.J.

Dobrushin, R. L. 1956a. Central limit theorems for non-stationary Markov chains I. Theor. Probability Appl. 1: 65-80.

Dobrushin, R. L. 1956b. Central limit theorems for non-stationary Markov chains II. Theor. Probability Appl. 1: 329-383.

Feller, W. 1968. An Introduction to Probability Theory and its Applications. Wiley, New York.

Griffeath, D. 1975. Uniform coupling of non-homogeneous Markov chains. J. Appl. Prob. 12:753-762.

Hajnal, J. 1956. The ergodic properties of non-homogeneous finite Markov chains. Proc. Camb. Phil. Soc. 52:66-67.

Huang, C. and Isaacson, D. 1976. Ergodicity using mean visit times. J. London Math. Soc. 14(2):570-576.

Isaacson, D. and Arnold, B. 1978. Strong ergodicity for continuous-time Markov chains. J. Appl. Prob. 15:699-706.

Isaacson, D. and Luecke, G. 1978. Strongly ergodic Markov chains and rates of convergence using spectral conditions. Stochastic Processes and Their Applications 7: 113-121.

Isaacson, D. and Madsen, R. 1976. Markov Chains Theory and Applications. Wiley, New York.

Johansen, S. and Ramsey F. A Bang-Bang representation for a 3x3 embeddable stochastic matrix. Preprint, Institute of Mathematical Statistics, University of Copenhagen, 1978.

Karlin, S. 1968. A First Course in Stochastic Processes. Academic Press, New York.

Parzen, E. 1962. Stochastic Processes. Holden-Day, San Francisco.

Pitman, J. W. 1974. Uniform rates of convergence for Markov chain transition probabilities. Z. Warh. Verw. Geb. 29:193-227.

Reuter, G. and Ledermann, W. 1953. On the differential equations for the transition probabilities of Markov processes with enumerably many states. Proc. Camb. Phil. Soc. 24:247-262.

Royden, H. 1968. Real Analysis. The Macmillan Company, Canada.

Rudin, W. 1973. Functional Analysis. McGraw-Hill, New York.

Tweedie, R. 1975. Sufficient conditions for regularity, recurrence and ergodicity of Markov processes. Math. Proc. Camb. Phil. Soc. 78:125-136.

Yong, P. L. 1976. Some results related to q-bounded Markov processes. Nanta Mathematica 8:34-41.

IX. ACKNOWLEDGMENTS

I would like to express my heartfelt thanks to my major professors. Professor Dean L. Isaacson and Professor Barry C. Arnold. Both individuals provided valuable insights and guidance with respect to every aspect of my graduate education at Iowa State University. It has been an honor and a pleasure to be associated with these two fine individuals on a professional and personal level.

I would like to express thanks to the rest of the members of my committee; Dr. David A. Harville, Dr. Glenn R. Luecke, and Dr. Edward Pollak.

I gratefully acknowledge the financial support of the Statistical Laboratory at Iowa State University and, in particular, Dr. Herbert A. David for giving me the opportunity to teach a varied number of courses.

I would like to thank Glen Meeden for being a friend.

I express a great deal of thanks to Dr. Thomas A. Starks for kindling my interest in statistics.

I would like to say thank you to the following group of my peers for making my stay at Iowa State University a worthwhile experience; Fred Dahm, Fred Dorn, Bill Sallas, Debra Shenk, Neil Werner, Michael Hand, Barry Simon, Richard Auer, Richard Dorsch, and Richard Welter.

I would like to thank my family, particularly my mother, for the encouragement and support she has provided over the years. A special thanks goes out to Paul and Suzanna Barcus.

Finally, I express appreciation to my wife, Irene, who has put up with my selfishness of time for the past two years.